Modelling data by the Choquet integral

Michel Grabisch*

LIP6, Université de Paris VI, Paris, France

1 Introduction

It is well known that the main tool for finding dependencies among data is the linear regression model, which expresses one or several variables in term of a linear combination of the others. Linear regression is based on the least square principle, and has been studied at length, its statistical performance are well known. In real situations however, the linear assumption happens to be often rather far from the reality, and models with low accuracy are produced. For this reason, more general models that offer more flexibility are looked for, the price to pay being that they are often much more complex to use.

In this paper, we are interested in using the Choquet integral [4] as a general non linear regression model. The Choquet integral is a generalization of the Lebesgue integral, defined with respect to a non classical measure, often called *fuzzy measure*, or *non-additive measure* or also *capacity*. When the underlying universe is finite, the Lebesgue integral reduces to a (convex) linear combination, hence can be assimilated to a particular class of regression models, where the coefficients are all positive and sum up to one. Hence, the Choquet integral offers a more general model, more precisely, as it will be seen below, it offers a set of (convex) linear models, each of them being defined in a simplex.

The Choquet integral has been successfully applied many times in classification [37,19,24], decision making under uncertainty [31,2,32], multicriteria decision making [8,35,25] and also data modelling [36]. The main difficulty is to determine efficiently the $2^n - 2$ coefficients of the model. This exponential complexity limits the applicability, although solutions to reduce this number to a polynomial size exist.

In this paper, we make a quick overview of the main works along this line, and describe recent works we performed. First sections are devoted to introduce the necessary material. In all the paper we consider a *single* regression problem, i.e. we want to model *one* variable y by some other variables $x_1, \dots, x_n$.

* On leave from THALES, Corporate Research Laboratory, Domaine de Corbeville, Orsay, France

2 The linear regression model

We recall briefly the linear regression model. Let y be a variable which we suppose we can explain or predict using a vector of variables $x = (x_1\ x_2\ \cdots\ x_n)^T$.

The general framework is rooted in estimation theory (where y is supposed to be unobservable, and thus has to be estimated), and based on the pioneering work of Gauss (see e.g.[1] for details). We consider x, y as random variables denoted X, Y. The assumption that y depends on x is that the distribution of $Y|X$ (posterior to observation) is different from the a priori distribution Y. Let us denote by $\hat{Y}(X)$ the estimated value of Y given an observation of X. The minimization of the variance of error $Y - \hat{Y}(X)$ leads to the unique solution $\hat{Y}(X) = E[Y|X]$, i.e. the conditional expectation. The linear hypothesis says that the estimated valued should be a linear expression of X, that is $\hat{Y}(X) = \alpha + \beta^T x$. Moreover, the expressions of α and β which minimize the variance of the error are given by:

$$\beta = \Gamma_{XY}\Gamma_Y^{-1} \tag{1}$$

$$\alpha = E[Y] - \beta E[X] \tag{2}$$

where Γ_{XY}, Γ_Y are the covariance matrices of X, Y and Y, i.e. $\Gamma_{XY} = E[(X - E[X])(Y - E[Y])^T]$, and $\Gamma_Y = E[(Y - E[Y])(Y - E[Y])^T]$. It turns out that the above linear model coincides with the (optimal) conditional expectation model when X, Y are gaussian.

The usual case of linear regression, where x, y are deterministic but the model is supposed to give only an approximation up to a random (model) error e is a particular case of the above linear estimation model:

$$y = \alpha + \beta^T x + e \tag{3}$$

Values of α, β which minimize the variance of the model error e follow directly from the above, when we have at disposal N data (realizations of x, y, denoted x^l, y^l, $l = 1, \ldots, N$), replacing expected values and covarianve matrices by the corresponding empirical expressions.

Usually, some quantities expressing the goodness of fit of the model to the data are computed.

3 Fuzzy measures and the Choquet integral

We present briefly necessary concepts around fuzzy measures and the Choquet integrals. Comprehensive treatments of this topic can be found in [5,18,17,30,38].

Let $N = \{1, \ldots, n\}$ be a finite set. A *capacity* or *fuzzy measure*, *non-additive measure* on N is any set function $\mu : \mathcal{P}(N) \longrightarrow \mathbb{R}^+$ such that $\mu(\emptyset) = 0$ and $A \subset B \subset N$ implies $\mu(A) \leq \mu(B)$ (monotonicity). μ is said to be *non-monotonic* if monotonicity does not hold.

The *conjugate* fuzzy measure of μ, denoted $\bar{\mu}$, is defined by $\bar{\mu}(A) = \mu(N) - \mu(A^c)$.

The Möbius transform of μ, denoted m^μ or m if there is no fear of ambiguity, is a set function on N defined by

$$m^\mu(A) := \sum_{B \subset A} (-1)^{|A|-|B|}\mu(B), \quad \forall A \subset N. \tag{4}$$

A fuzzy measure is said to be *additive* if $\mu(A \cup B) = \mu(A) + \mu(B)$ whenever $A \cap B = \emptyset$. A *k-additive measure* is a fuzzy measure such that $m(A) = 0$ for all $A \subset N$ such that $|A| > k$, and there is at least one A of k elements such that $m(A) \neq 0$. 1-additivity coincides with additivity.

An important notion for the interpretation of fuzzy measures is the one of Shapley and interaction indices. For any $i \in N$, the Shapley index [33] of i is defined by:

$$\phi_i := \sum_{K \subset N \setminus i} \frac{(n - |K| - 1)!|K|!}{n!}[\mu(K \cup \{i\}) - \mu(K)]. \tag{5}$$

The Shapley index satisfies $\sum_{i=1}^n \phi_i = \mu(N)$, and can be interpreted as the overall importance of i. The concept of interaction for a pair of elements $i, j \in N$ has been proposed by Murofushi and Soneda [28].

$$I_{ij} := \sum_{K \subset N \setminus \{i,j\}} \frac{(n - |K| - 2)!|K|!}{(n-1)!}[\mu(K \cup \{i,j\}) - \mu(K \cup \{i\}) - \mu(K \cup \{j\}) + \mu(K)]. \tag{6}$$

It represents a kind of (positive or negative) synergy between elements. The definition has been extended by the author to any number of elements [11]:

$$I(A) := \sum_{K \subset N \setminus A} \frac{(n - |K| - |A|)!|K|!}{(n - |A| + 1)!} \sum_{B \subset A} (-1)^{|A|-|B|}\mu(K \cup B), \forall A \subset N. \tag{7}$$

Note that $I(\{i\}) = \phi_i$, and $I(\{i,j\}) = I_{ij}$. Also, it is easy to show that for an additive measure, $I(A) = 0$ whenever $|A| > 1$, and $\phi_i = \mu(\{i\})$. More generally, for a k-additive measure, $I(A) = 0$ whenever $|A| > k$.

We introduce now the (discrete) Choquet integral on N. We assimilate (positive) real-valued functions on N to points in $\mathbb{R}^n_+$. Let $x \in \mathbb{R}^n_+$, with components $x_i, i = 1, \ldots, n$, and μ be a fuzzy measure. The *Choquet integral* of x w.r.t. μ is defined by:

$$\mathcal{C}_\mu(x) := \sum_{i=1}^n x_{\sigma(i)}[\mu(A_{\sigma(i)}) - \mu(A_{\sigma(i+1)})] \tag{8}$$

where σ is a permutation of the elements of N such that $x_{\sigma(1)} \leq \cdots \leq x_{\sigma(n)}$, $A_{\sigma(i)} := \{\sigma(i), \sigma(i+1), \ldots, \sigma(n)\}$, and $A_{\sigma(n+1)} := \emptyset$. Remark that for a given permutation σ, the region $\{x \in \mathbb{R}^n_+ | x_{\sigma(1)} \leq \cdots \leq x_{\sigma(n)}\}$ is a simplex. We call *canonical simplexes* the set of all such simplexes, considering all possible permutations on N. They form a partition of $\mathbb{R}^n_+$. Moreover, continuity is ensured over all canonical simplexes.

4 The Choquet integral regression model

The regression based on Choquet integral is a generalization of Eq. (3) in the following sense:

$$y = \alpha + \mathcal{C}_\mu(x) + e$$

using previous notations, where μ is a (non monotonic in general) fuzzy measure. From Section 3, we know that more explicitely the model writes:

$$y = \alpha + \sum_{i=1}^{n} w_{\sigma(i)} x_{\sigma(i)} + e, \tag{9}$$

with $w_{\sigma(i)} = \mu(A_{\sigma(i)}) - \mu(A_{\sigma(i+1)})$. Observe that in any case $\sum_{i=1}^n w_{\sigma(i)} = \mu(N) = 1$, but the $w_{\sigma(i)}$'s are positive (i.e. we get a convex sum) iff μ is monotone.

Although this model is clearly more general than the linear one, some restrictions have to be pointed out. The first one is that the definition of the Choquet integral (8) is given for positive integrands, hence x should belong to $\mathbb{R}^n_+$. For real-valued integrands, two definitions exist, the symmetric and the asymmetric ones, which coincide when the fuzzy measure is additive, i.e. when the Choquet integral model collapses to the linear one. As a consequence, when $x \in \mathbb{R}^n$, two models are possible. The expressions of the symmetric Choquet integral $\check{\mathcal{C}}_\mu$ (called also Šipoš integral) and asymmetric Choquet integral $\mathcal{C}_\mu$ are:

$$\check{\mathcal{C}}_\mu(x) = \mathcal{C}_\mu(x^+) - \mathcal{C}_\mu(x^-)$$
$$\mathcal{C}_\mu(x) = \mathcal{C}_\mu(x^+) - \mathcal{C}_{\overline{\mu}}(x^-)$$

where $x_i^+ = x_i \vee 0$, and $x_i^- = -x_i \vee 0$, for all i.

The second point is more of importance, and it is rooted in the definition of the Choquet integral. Considering a permutation σ on N, the corresponding canonical simplex is the locus of the points satisfying $x_{\sigma(1)} \leq x_{\sigma(2)} \cdots \leq x_{\sigma(n)}$. These simplexes are central in the definition of the Choquet integral, but they suppose that one can meaningfully compare the variables x_i's ! In real applications, variables are most often expressed with different units, e.g. in medicine one try to model the blood pressure in terms of age, height and weight. Clearly, these three variables are not commensurate, so applying the Choquet integral here is meaningless.

To the opinion of the author, the last above mentionned point prevents the Choquet integral to be used as a general non linear regression model, in any situation, despite the fact that this kind of model is advocated by some authors as Wang *et al.* [41,40]. However, this drawback disappears if one considers only commensurate variables. This is always the case in multicriteria evaluation problems, and multi-attribute classification, a domain where the Choquet integral has been successfully applied many times (see Section 6 for some examples).

In multicriteria evaluation, N is the set of criteria or attributes, and x_i is not the value taken by attribute i for some object, but represents the *satisfaction degree* or *attractiveness* felt by the decision maker in view of the value taken by attribute i. Depending on the application and the precise meaning attached to the x_i's, the underlying scale could be bounded unipolar (e.g. $[0,1]$), unipolar (e.g. $\mathbb{R}^+$) or bipolar (e.g. $\mathbb{R}$). In this last case, the scale could be a difference scale or a ratio scale, which determines the kind of integral to be used (symmetric or asymmetric): see details in [16].

In classification, N is the set of attributes, and x_i represents the membership degree of a given object to a given class, knowing only the value of attribute i. The membership degree is a bounded unipolar concept, so that the underlying scale is $[0,1]$.

In these two domains of application, the fuzzy measure is asked to be monotone, since this entails the monotonicity of the model, a natural requirement in these two fields.

We end this section by giving some words on the advantage of such models. Coefficients in a linear regression model are easy to interpret, as they represent the “weight” of a given variable in the model. In the case of the Choquet integral model, there are too many weights, and these weights live only in some canonical simplex, so that at first sight, the model is not easily interpretable. However, an interpretation in terms of the Shapley value and interaction indices permits to have a clear view of the model. In fact, the merit of the Choquet integral is to bring a powerful tool to model interaction between variables (see an explanation of interaction in e.g. [8,10], see also a particularly simple geometrical interpretation when $n = 2$ in [13]), which is theoretically well founded (see an axiomatization in the spirit of the Shapley value in [21]).

5 Determining the coefficients of the model

In the case of the linear regression, a unique solution was obtained, based on results of estimation theory. In the case of Choquet integral, no such result is available, and the coefficients of the model (i.e. the fuzzy measure) are obtained through an optimization procedure, whose solution is not unique in general. As in the linear regression model, one wants to minimize the squared error, that is:

$$E = \sum_{l=1}^{N} [y^l - \mathcal{C}_\mu(x^l)]^2 \tag{10}$$

under constraints if the fuzzy measure is asked to be monotone. If no constraint (apart positiveness) exists, then the problem is not so much difficult to solve, since it reduces to a usual least square problem. Indeed, it can be shown (see e.g. [18,20], and originally [27]) that E can be expressed in a quadratic form:

$$\frac{1}{2}\mathbf{u}^T\mathbf{D}\mathbf{u} + \mathbf{c}^T\mathbf{u}$$

where $\mathbf{u}$ is a $(2^n - 2)$ dimensional vector containing all the coefficients of the fuzzy measure μ (except $\mu(\emptyset)$ and $\mu(N)$ which are fixed), $\mathbf{D}$ is a $(2^n - 2)$ dimensional square matrix, and $\mathbf{c}$ a $(2^n - 2)$ dimensional vector. The constraints of monotonicity can be expressed with $\mathbf{u}$ under a linear form:

$$\mathbf{A}\mathbf{u} + \mathbf{b} \geq \mathbf{0}$$

where $\mathbf{A}$ is a $n(2^{n-1} - 1) \times (2^n - 2)$ matrix, and $\mathbf{b}$ a $n(2^{n-1} - 1)$ dimensional vector. Thus we obtain a quadratic program, which can be solved using standard techniques. In [26], Miranda and Grabisch study at length the properties of this quadratic program. There is no unique solution in general, and one important question is to know the minimum number of data required in order to have a "good" solution. This point remains however not completely clear.

Although the preceding approach provides an optimal solution, it happens that in some cases (large n, few data,. . .) the problem becomes ill-conditioned and bad results occur. Also, in practical applications, the optimal solution obtained does not always satisfy the decision maker, giving "extreme" values (near 0 and 1) and far from the equilibrium point $\mu(A) = 1/|A|$ for all $A \subset N$. For this reason, the author has proposed a suboptimal algorithm, called HLMS (Heuristic Least Mean Squares), based on the gradient algorithm and the idea of equilibrium point [7]. This algorithm gives an error very near the optimal one, while being much less memory and time consuming.

A third way to solve the optimization problem is to use heuristic algorithms, such as genetic algorithms (GA). Many authors have proposed methods based on GA to determine the fuzzy measure, although most of them are restricted to λ-measures (see e.g. [3]). A typical approach to determine (general) fuzzy measures is the one presented by Wang et al. in [39]. We describe in the whole the method, and propose some improvement we performed.

- the encoding of μ is done as follows: $\mu(A)$ is coded in a gene, for all $A \subset N$, $A \neq \emptyset, N$, so that a chromosome has $2^n - 2$ genes coding a given μ. Each gene is coded as a binary number, using p bits.
- the population of chromosomes is between 100 and 1000, and is randomly generated (uniform numbers generated on $[0, 1]$, plus a test for monotonicity).
- the fitness of a chromosome is defined by $\frac{1}{1+E}$, where E is the above defined quadratic error.

- the probability of choosing parents is proportional to the fitness value. Then reproduction is done according to three different processes: two-point crossover, three-bit mutation, and two-point realignment (see [39] for details).
- the population size is kept constant by always selecting the best individuals by their fitness.

Although this is not explained in the paper, we suppose that a test of monotonicity is performed over new chromosomes, and those which do not satisfy the conditions are eliminated.

We proposed a more general version, allowing to handle k-additive measures, and improving the optimization process in order to reduce the learning time and improve performance.

1. **General features:** the program accepts several error criteria (sum of squares, of absolute values, etc.), and several fuzzy integrals (Choquet, Šipoš Sugeno, etc.).
2. **Coding:** it is the same than Wang's approach, at the difference that we code the Möbius transform when dealing with k-additive measures. In this last case, the number of genes to code is only $\sum_{l=1}^{k}\binom{n}{l} - 1$.
3. **Selecting the genes:** recall that for a given datum x is associated a permutation σ_x on N ordering the components of x in increasing order, and to each permutation σ is associated a maximal chain C_σ in the lattice of values of μ: only the coefficients belonging to the chain are used in the computation of the integral. Henceforth, for a given set of learning data $\mathcal{X}$, the coefficients that will be effectively used in the computations form the set $\bigcup_{x\in\mathcal{X}} C_{\sigma_x}$, which may be much smaller than the whole lattice. In the learning stage, only these genes will be modified by reproduction. Note that this procedure works only for the case where μ is encoded, not its Möbius transform.
4. **reproduction, test of monotonicity and selection:** these operations are performed on the current population, unless some stagnation is observed, in which case the former population of best individuals is chosen.
   - There are several modes of reproduction which can be chosen, either producing two children or one child, using one point crossover or two points crossover, selecting parents either at random, or neighbours, or according to a probability depending on the fitness function, and using mutation or not.
   - the test of monotonicity depends on the type of information which is coded (either μ or its Möbius transform). In the latter, the standard test of monotonicity for the Möbius transform is performed. In the former case, recall that only genes in $\bigcup_{x\in\mathcal{X}} C_{\sigma_x}$ (called *significant*) are modified. Non significant genes are nevertheless updated in order to satisfy monotonicity of μ, in a spirit similar as in [7]. Specifically, in the lattice formed by μ, non significant genes are given the maximum value of lower neighbours in the lattice (i.e. the smallest one), and

the test of monotonicity for significant genes is to verify that their value is at least as large as all its lower neighbours.
- the selection is done in the usual way, keeping the best individuals at constant population size.

5. **Initial population:** random generation of monotone fuzzy measures or Möbius transforms.

We have compared our approach with the algorithm of Wang on a small example ($n = 4$) given in [39]. Wang *et al.* found an error of 0.0088 in 2 minutes on a 120 MHz processor, while we found an error of 0.0001 in less than 5 seconds on a 733 MHz processor. This shows clearly the improvement.

Tests comparing on various data sets the GA approach with the above described quadratic programming approach show that, as expected, the latter outperforms the former, in precision (both can give very close results when n is low), but mainly in learning time. In our opinion, GA's are not useful for the Choquet integral, since classical optimal methods can do the job very efficiently, but they become useful for highly non linear functionals as the Sugeno integral. A careful study devoted to the Sugeno integral has yet to be done.

6 Related works and examples

This section gathers various related works, using Choquet integral as a basic tool for modelling.

6.1 Classification

The Choquet integral has been used as an information fusion tool for performing multi-sensor classification, each sensor being able to give a classification. The first attempt in this direction was done by Tahani and Keller [37], using a λ-measure. At the same time, Grabisch and Sugeno proposed a more general approach for classification, similar to the Bayesian one [22,23].

Basically, the idea is to describe each class by a set of typical fuzzy sets or possibility distribution (one per attribute), and to define a fuzzy measure per class, on the set of attributes. For class j, the fuzzy measure μ_j expresses to what degree coalitions of attributes are able to distinguish class j from the others. The method has been described at length in several publications, we refer the reader to the following literature: [19,9,12]. See also [18,24] for an overview of related works in classification and image processing.

6.2 Subjective evaluation

As said in Section 4, multicriteria evaluation is particularly suited to the use of a Choquet integral model (or similar integrals as the Sugeno integral.

A precise distinction of these two types of integrals is however still a topic of research). The term “subjective” refers to the fact that most of concepts modelled so far are highly subjective, so that no clear mathematical model can be derived simply from analysis of the problem, as the following examples will made it clear. It is worth noting that the first application of fuzzy measure was indeed evaluation, of woman faces (how subjective!), by Sugeno in his thesis [34].

Many applications in this field have been done in Japan during the eighties, such as modelling opinion on nuclear energy [29], evaluating the quality of printed color images, speakers, and so on (see an overview of these applications in [18]), rather in an ad hoc way. Later, using the concept of interaction as a basic tool for analysis, and k-additive measures, the author has performed several applications, as the evaluation of richness of a cosmetic [14], of discomfort in a seated position during a long time [15], of mental work load when performing some task, etc.

6.3 The model of Kwon and Sugeno

In [35,25], Kwon and Sugeno propose a model of multicriteria evaluation based on Choquet integral, but in a rather different approach as the one we presented above. Considering a set N of n criteria, their argument is that, except for low values of n, a model based on fuzzy measures becomes intractable due to the exponential complexity (2^n coefficients). Based on works of Fujimoto [6] about inclusion-exclusion coverings, they propose to replace the Choquet integral defined with respect to a fuzzy measure μ on N by a sum of p Choquet integrals w.r.t. fuzzy measures $\mu_1, \ldots, \mu_p$ on subsets $C_1, \ldots, C_p$ of N, such that $\bigcup_{i=1}^p C_i = N$, i.e. $\{C_1, \ldots, C_p\}$ is a covering of N. In equation:

$$y = \sum_{i=1}^{p} \mathcal{C}_{\mu_i}(x) + e. \tag{11}$$

In the above, x in $\mathcal{C}_{\mu_i}(x)$ is of course restricted to C_i. The μ_i's are non monotonic fuzzy measures, and e is a modelling error supposed to be a zero mean Gaussian random variable with variance σ^2 (denoted $\mathcal{N}(0, \sigma^2)$). x and y represents satisfaction degrees, and are supposed to be commensurable.

The identification of the parameters of the model is done as follows. Considering N independent realizations of y with data $x^1, \ldots, x^N$, denoted $y^1, \ldots, y^N$, the joint distribution of $y^1, \ldots, y^N$ is $\prod_{l=1}^N \mathcal{N}(\sum_{i=1}^p \mathcal{C}_{\mu_i}(x^l), \sigma^2)$, and assuming that the covering $\{C_1, \ldots, C_p\}$ is known, the μ_i's are determined in order to minimize the residual variance σ^2 of the error, which amounts to the squared error criterion presented above:

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{l=1}^{N} \left[y^l - \sum_{i=1}^{p} \mathcal{C}_{\mu_i}(x^l) \right]$$

Since the μ_i's are not restricted to be monotone, the estimation can be done with the least square method.

The second step is to determine the optimal covering. Here genetic algorithms are used. A particular covering is coded in a chromosome containing $2^n - 1$ genes corresponding to all subsets of N, except the empty set. The gene corresponding to a particular subset is set to 1 if the covering contains this subset, and to 0 otherwise. We refer the reader to [35,25] for further details of implementation, and detail only the fitness function. They use the Bayesian Information Criterion (BIC) defined as:

$$\text{BIC} = N \log \hat{\sigma}^2 + k \log N$$

where k is the number of independent parameters in the model. The criterion expresses the fact that a balance should be kept between precision of the model (low $\hat{\sigma}^2$) and simplicity (low k).

We make some comments about this model.

- Although equation (11) seems to be very general, it coincides exactly with a Choquet integral whenever the covering is an inclusion-exclusion covering (IEC) (see [6]). If not, it is more general, but properties of this class of operators have not been studied by the authors. The main interest of this way of decomposing the integral has the advantage of exhibiting some structure in the evaluation model and to lower complexity. Another way to do this is to use k-additive measures. In fact, Fujimoto and Murofushi have shown that an IEC corresponds more or less to the subsets where the Möbius transform is non zero (more exactly, the finest irreducible (i.e. non redundant) IEC is the set of all maximal subsets of non-zero Möbius transform).
- The use of non-monotonic measures, although more general, is questionable since we are here in an evaluation process, where x is a vector of satisfaction degree. It is a fundamental assumption in such context that the model is monotonic, i.e. the increase of a satisfaction degree on some criterion cannot decrease the overall satisfaction degree y. But this condition implies the monotonicity of the fuzzy measure.
- Results presented in [25] on real data (evaluation of motorcycles, according to different categories of population) show that the performance of the model are very close to a usual linear model (almost same error, sometimes better for the linear model, but with a slightly better BIC value). A much better modelling error is obtained using a (general) Choquet integral, at the price of having much more coefficients. Also, the subsets in the covering are almost reduced to singletons and sometimes pairs, which explains why the performance is so close to a linear model. This may be due to the fitness function since it is linear in k and only logarithmic in $\hat{\sigma}^2$.

6.4 Discovering links betweeen variables

Lastly we present briefly a general model of regression proposed by Wang *et al.* [40]. Supposing to have at disposal data about some variables, the algorithm tries to find the best dependencies among the variables, and each sub-model is given by a Choquet integral regression model.

Specifically, dependencies among variables are expressed as an oriented acyclic graph. Each node y with entering arrows can be explained by the parent nodes $x_1, \ldots, x_n$ (corresponding to the entering arrows), and this forms a sub-model. Each sub-model is expressed through a Choquet integral:

$$y = q\,\mathcal{C}_\mu + e$$

where q is a real constant (positive?), μ a fuzzy measure, and e is the modelling error, supposed to be Gaussian and centered on a value c.

The basic idea is the following: the best acyclic graph is obtained by genetic algorithms, while for each sub-model, an algorithm similar to the one proposed in [7] is used, and q, c are estimated as in a linear regression model.

To the opinion of the author, the approach, although very general and powerful, does not avoid the restriction we have indicated in Section 4. First of all, one cannot combine with a Choquet integral variables which are not commensurate, hence the above model cannot be used in data mining in general as claimed by the authors. Second, the fact that fuzzy measures are used (and not non monotonic fuzzy measures) implies that each sub-model is monotonically increasing w.r.t each variable, an assumption which is again restrictive.

7 Conclusion

In this paper, we have presented an overview of methods using the Choquet integral as a tool of modelling data. The main advantage compared to linear models is that they are able to take into account interactions and dependency between variables. We have also presented the limitations of such approaches, which to our sense, should be limited to the modelling of commensurate variables, a situation which is typically encountered in multicriteria evaluation, multi-attribute classification (after all, a Choquet integral is not more than a generalization of expected value. . .).

We have also presented a new approach of determination of fuzzy measures by genetic algorithms. This could be used for other fuzzy integrals, such as the Sugeno integral, where classical optimization methods fail.

8 Acknowledgment

The author is grateful to Ch. de Rivière and J. Wasong for their active participation in the project of determining fuzzy measures by genetic algorithms.

References

1. B.D.O. Anderson and J.B. Moore. *Optimal filtering.* Prentice Hall, 1979.
2. A. Chateauneuf. Modeling attitudes towards uncertainty and risk through the use of Choquet integral. *Annals of Operations Research*, 52:3–20, 1994.
3. T.Y. Chen and J.C. Wang. Identification of λ-measures using sampling design and genetic algorithms. *Fuzzy Sets and Systems*, 123:321–341, 2001.
4. G. Choquet. Theory of capacities. *Annales de l'Institut Fourier*, 5:131–295, 1953.
5. D. Denneberg. *Non-Additive Measure and Integral.* Kluwer Academic, 1994.
6. K. Fujimoto and T. Murofushi. Hierarchical decomposition of the Choquet integral. In M. Grabisch, T. Murofushi, and M. Sugeno, editors, *Fuzzy Measures and Integrals — Theory and Applications*, pages 94–104. Physica Verlag, 2000.
7. M. Grabisch. A new algorithm for identifying fuzzy measures and its application to pattern recognition. In *Int. Joint Conf. of the 4th IEEE Int. Conf. on Fuzzy Systems and the 2nd Int. Fuzzy Engineering Symposium*, pages 145–150, Yokohama, Japan, March 1995.
8. M. Grabisch. The application of fuzzy integrals in multicriteria decision making. *European J. of Operational Research*, 89:445–456, 1996.
9. M. Grabisch. The representation of importance and interaction of features by fuzzy measures. *Pattern Recognition Letters*, 17:567–575, 1996.
10. M. Grabisch. Alternative representations of discrete fuzzy measures for decision making. *Int. J. of Uncertainty, Fuzziness, and Knowledge Based Systems*, 5:587–607, 1997.
11. M. Grabisch. k-order additive discrete fuzzy measures and their representation. *Fuzzy Sets and Systems*, 92:167–189, 1997.
12. M. Grabisch. Fuzzy integral for classification and feature extraction. In M. Grabisch, T. Murofushi, and M. Sugeno, editors, *Fuzzy Measures and Integrals — Theory and Applications*, pages 415–434. Physica Verlag, 2000.
13. M. Grabisch. A graphical interpretation of the choquet integral. *IEEE Tr. on Fuzzy Systems*, 8:627–631, 2000.
14. M. Grabisch, J.M. Baret, and M. Larnicol. Analysis of interaction between criteria by fuzzy measure and its application to cosmetics. In *Int. Conf. on Methods and Applications of Multicriteria Decision Making*, pages 22–25, Mons, Belgium, May 1997.
15. M. Grabisch, J. Duchêne, F. Lino, and P. Perny. Subjective evaluation of discomfort in sitting position. *Fuzzy Optimization and Decision Making*, submitted.
16. M. Grabisch and Ch. Labreuche. To be symmetric or asymmetric? A dilemna in decision making. In J. Fodor, B. De Baets, and P. Perny, editors, *Preferences and Decisions under Incomplete Knowledge*, pages 179–194. Physica Verlag, 2000.
17. M. Grabisch, T. Murofushi, and M. Sugeno. *Fuzzy Measures and Integrals. Theory and Applications (edited volume).* Studies in Fuzziness. Physica Verlag, 2000.
18. M. Grabisch, H.T. Nguyen, and E.A. Walker. *Fundamentals of Uncertainty Calculi, with Applications to Fuzzy Inference.* Kluwer Academic, 1995.
19. M. Grabisch and J.M. Nicolas. On the performance of classification techniques based on fuzzy integrals. In *5th Int. Fuzzy Systems Assoc. Congress*, pages 163–166, Seoul, Korea, July 1993.

20. M. Grabisch and J.M. Nicolas. Classification by fuzzy integral — performance and tests. *Fuzzy Sets & Systems, Special Issue on Pattern Recognition*, 65:255–271, 1994.
21. M. Grabisch and M. Roubens. An axiomatic approach to the concept of interaction among players in cooperative games. *Int. Journal of Game Theory*, 28:547–565, 1999.
22. M. Grabisch and M. Sugeno. Fuzzy integral with respect to dual measures and its application to multi-attribute pattern recognition. In *6th Fuzzy Systems Symposium*, pages 205–209, Tokyo, Japan, September 1990. in japanese.
23. M. Grabisch and M. Sugeno. Multi-attribute classification using fuzzy integral. In *1st IEEE Int. Conf. on Fuzzy Systems*, pages 47–54, San Diego, CA, March 1992.
24. J.M. Keller, P.D. Gader, and A.K. Hocaoğlu. Fuzzy integrals in image processing and recognition. In M. Grabisch, T. Murofushi, and M. Sugeno, editors, *Fuzzy Measures and Integrals — Theory and Applications*, pages 435–466. Physica Verlag, 2000.
25. S.H. Kwon and M. Sugeno. A hierarchical subjective evaluation model using non-monotonic fuzzy measures and the Choquet integral. In M. Grabisch, T. Murofushi, and M. Sugeno, editors, *Fuzzy Measures and Integrals — Theory and Applications*, pages 375–391. Physica Verlag, 2000.
26. P. Miranda and M. Grabisch. Optimization issues for fuzzy measures. *Int. J. of Uncertainty, Fuzziness, and Knowledge-Based Systems*, 7(6):545–560, 1999.
27. T. Mori and T. Murofushi. An analysis of evaluation model using fuzzy measure and the Choquet integral. In *5th Fuzzy System Symposium*, pages 207–212, Kobe, Japan, 1989. In Japanese.
28. T. Murofushi and S. Soneda. Techniques for reading fuzzy measures (III): interaction index. In *9th Fuzzy System Symposium*, pages 693–696, Sapporo, Japan, May 1993. In Japanese.
29. T. Onisawa, M. Sugeno, Y. Nishiwaki, H. Kawai, and Y. Harima. Fuzzy measure analysis of public attitude towards the use of nuclear energy. *Fuzzy Sets & Systems*, 20:259–289, 1986.
30. E. Pap. *Null-Additive Set Functions*. Kluwer Academic, 1995.
31. D. Schmeidler. Subjective probability and expected utility without additivity. *Econometrica*, 57(3):571–587, 1989.
32. U. Schmidt. *Axiomatic Utility Theory under Risk*. Number 461 in Lectures Notes in Economics and Mathematical Systems. Springer Verlag, 1998.
33. 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*, number 28 in Annals of Mathematics Studies, pages 307–317. Princeton University Press, 1953.
34. M. Sugeno. *Theory of fuzzy integrals and its applications*. PhD thesis, Tokyo Institute of Technology, 1974.
35. M. Sugeno and S.H. Kwon. A clusterwise regression-type model for subjective evaluation. *J. of Japan Society for Fuzzy Theory and Systems*, 7(2):291–310, 1995.
36. M. Sugeno and S.H. Kwon. A new approach to time series modeling with fuzzy measures and the Choquet integral. In *Int. Joint Conf. of the 4th IEEE Int. Conf. on Fuzzy Systems and the 2nd Int. Fuzzy Engineering Symp.*, pages 799–804, Yokohama, Japan, March 1995.

37. H. Tahani and J.M. Keller. Information fusion in computer vision using the fuzzy integral. *IEEE Tr. on Systems, Man, and Cybernetics*, 20(3):733–741, 1990.
38. Z. Wang and G.J. Klir. *Fuzzy measure theory.* Plenum, 1992.
39. Z. Wang, K.S. Leung, and J. Wang. A genetic algorithm for determining nonadditive set functions in information fusion. *Fuzzy Sets and Systems*, 102:462–469, 1999.
40. Z. Wang, K.S. Leung, M.L. Wong, J. Fang, and K. Xu. Nonlinear nonnegative multiregressions based on Choquet integrals. *Int. J.of Approximate Reasoning*, 25:71–87, 2000.
41. K. Xu, Z. Wang, and K.S. Leung. Using a new type of nonlinear integral for multiregression: an application of evolutionary algorithms in data mining. In *Proc. of IEEE Int. Conf. on Systems, Man and Cybernetics*, pages 2326–2331, 1998.