Article

Forming a Hierarchical Choquet Integral with a GA-Based Heuristic Least Square Method

Chin-Yi Chen ¹ and Jih-Jeng Huang ^{2,*}

¹ Department of Business Administration, Chung Yuan Christian University, Taoyuan 32023, Taiwan; iris@cycu.edu.tw

² Department of Computer Science & Information Management, Soochow University, No.56 Kueiyang Street, Section 1, Taipei 100, Taiwan

* Correspondence: jjhuang@scu.edu.tw; Tel.: +886-02-23111531 (ext. 3813)

Received: 14 October 2019; Accepted: 23 November 2019; Published: 1 December 2019

Abstract: Identifying the fuzzy measures of the Choquet integral model is an important component in resolving complicated multi-criteria decision-making (MCDM) problems. Previous papers solved the above problem by using various mathematical programming models and regression-based methods. However, when considering complicated MCDM problems (e.g., 10 criteria), the presence of too many parameters might result in unavailable or inconsistent solutions. While k -additive or p -symmetric measures are provided to reduce the number of fuzzy measures, they cannot prevent the problem of identifying the fuzzy measures in a high-dimension situation. Therefore, Sugeno and his colleagues proposed a hierarchical Choquet integral model to overcome the problem, but it required the partition information of the criteria, which usually cannot be obtained in practice. In this paper, we proposed a GA-based heuristic least mean-squares algorithm (HLMS) to construct the hierarchical Choquet integral and overcame the above problems. The genetic algorithm (GA) was used here to determine the input variables of the sub-Choquet integrals automatically, according to the objective of the mean square error (MSE), and calculated the fuzzy measures with the HLMS. Then, we summed these sub-Choquet integrals into the final Choquet integral for the purpose of regression or classification. In addition, we tested our method with four datasets and compared these results with the conventional Choquet integral, logit model, and neural network. On the basis of the results, the proposed model was competitive with respect to other models.

Keywords: fuzzy measure; hierarchical Choquet integral; genetic algorithm; heuristic least mean-squares algorithm (HLMS)

1. Introduction and Presentation of the Problem

The fuzzy integral is applied to evaluate multi-criteria decision-making (MCDM) and consider the non-additive problems between criteria [1,2]. Two well-known fuzzy integral models are widely used in the field of MCDM: The Sugeno and the Choquet integral models. Reference [3] provided the concept of the λ —measure to replace the additive requirement of classical measures, which require the properties of monotonicity and continuity. Many studies have successfully employed the Sugeno integral model in various fields, for example, data classification [4], face recognition [5], and qualitative data analysis [6]. However, these decomposable fuzzy measures cannot be super-additive for some subsets of criteria and can be sub-additive for other subsets [1]. Hence, the decomposable coefficients can only express either sub-additive or super-additive measures for the whole set of criteria and restrict the Sugeno integral model from fitting into particular MCDM problems.

However, the Choquet integral model was proposed by Choquet in 1954 [7] to represent certain kinds of interactions among criteria using the concepts of redundancy and support/synergy, which are

related to the Shapley value in cooperative games and are explained in References [8,9]. The problem of identifying fuzzy measures in the Choquet integral can be considered as a mathematical programming problem to reduce the error between the predicted value and evaluated output, for example, the maximum split [10], minimum variance [11], and less constrained approach [12]. Furthermore, the heuristic least mean-squares algorithm (HLMS) was also proposed in Reference [13] to identify fuzzy measures by learning from data based on the gradient descent algorithm. However, it is hard to identify $2^n - 2$ fuzzy measures for a huge n criteria situation in practice.

Hence, Reference [14] proposed the concept of k -additive measures and Reference [15] proposed p -symmetric measures to reduce the number of coefficients to $\sum_{j=1}^k \binom{n}{j}$ and $[(|A_1|+1) \times \cdots \times (|A_p|+1)] - 2$, where $\{A_1, \dots, A_p\}$ denotes the partition of criteria. While k -additive measures significantly reduce the complexity of identifying coefficients, determining the appropriate coefficients still overloads the ability of a decision-maker in solving practical problems (e.g., for $n = 10$ and $k = 2$, 55 coefficients should be identified). In addition, information on the partition of criteria is usually unavailable in practice.

In order to overcome the practical problem of identifying fuzzy measures in the Choquet integral model, Sugeno and his colleagues proposed the hierarchical Choquet integral models [16] to decompose the Choquet integral model into several sub-Choquet integral models based on information from the inclusion–exclusion covering (IEC) and inter-additive partition (IAP). While the hierarchical Choquet integral models significantly reduce the estimated parameters, information on the IEC or IAP is not always available. In addition, the hierarchical decomposition theorem above cannot consider possible errors when identifying fuzzy measures and might be unsuitable for use with real data.

In this paper, we use a genetic algorithm (GA) to construct hierarchical Choquet integral models to reduce the estimated fuzzy measures without information on the IEC or IAP and derive the appropriate fuzzy measures based on the mean square error between the predicted values and the ground truth data. In addition, we extend the proposed method to consider the tasks of regression and classification. Four datasets are used to compare the results between the proposed method and the conventional Choquet integral, logit model, and neural network. The experimental results show that the proposed method is competitive with respect to the criteria of the mean square error and accuracy.

2. The Choquet Integral

The Choquet integral was proposed by Choquet [7] to generalize the weighted arithmetic mean to consider complementarity or substitutivity among criteria. In this section, we give a short introduction about the theory of the fuzzy measure and the Choquet integral [1,13,17,18] as follows.

Definition 1. Let X be a measurable set that is endowed with the properties of algebra, where $\mathfrak{N}$ is all the subsets of X . A fuzzy measure μ defined on the measurable space $(X, \mathfrak{N})$ is a set functions $\mu : \mathfrak{N} \rightarrow [0, 1]$, which satisfy the following properties:

- (1). $\mu(\emptyset) = 0$, $\mu(X) = 1$;
- (2). For all $A, B \in \mathfrak{N}$, if $A \subseteq B$ then $\mu(A) \leq \mu(B)$ (monotonicity).

Definition 2. Let $(X, \mathfrak{N}, \mu)$ be a fuzzy measure space and f be a nonnegative measurable function. The Choquet integral of the nonnegative simple function $f : X \rightarrow \mathbb{R}^+$ with respect to the fuzzy measure $\mu : \mathfrak{N} \rightarrow [0, 1]$ can be defined as [19]

$$\int f \cdot d\mu := \sum_{i=1}^n [f(x_i) - f(x_{i-1})] \mu(A_i), \quad (1)$$

where i indicates a permutation of the set $\{x_1, \dots, x_n\}$ such that $0 \leq f(x_1) \leq f(x_2) \leq \cdots \leq f(x_n)$ and $A_i := \{x_i, \dots, x_n\}$.

Since the fuzzy capacities are hard to be estimated under large criteria, many methods have been proposed to determine the fuzzy capacities without considering $2^n - 2$ fuzzy capacities, e.g., the Möbuis transformation or k -additive Choquet integral.

Definition 3. The Möbuis transformation of μ is a set function on N defined by:

$$m_\mu(T) := \sum_{B \subseteq T} (-1)^{|T|-|B|} \mu(B), \quad \forall T \subseteq N. \quad (2)$$

However, any set of 2^n coefficients does not necessarily correspond to the Möbuis transformation of a capacity on N . Hence, the boundary and monotonicity conditions must be ensured as follows [20]:

$$\begin{cases} m(\emptyset) = 0, & \sum_{T \subseteq N} m(T) = 1, \\ \sum_{i \in B \subseteq T} m(B) \geq 0, & \forall T \subseteq N, \quad \forall i \in T. \end{cases} \quad (3)$$

Then, the Choquet integral of x with respect to μ is given by:

$$C_{m_\mu}(x) = \sum_{T \subseteq N} m_\mu(T) \Lambda x_i, \quad (4)$$

where the symbol Λ denotes the minimum operator.

Definition 4. The Shapley index of i is defined by:

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

The Shapley index can be interpreted as the overall importance of i . We can also consider the concept of interaction for a pair of elements $i, j \in N$ from Murofushi and Soneda (1993) as follows:

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

The above index can be interpreted as the positive or negative synergy between the elements.

Definition 5. A fuzzy measure μ is said to be k -additive if its Möbuis transformation satisfies $m(A) = 0$ for any $A \subseteq X$ such that $|A| > k$, and at least one subset A of X exactly k elements exists, such that $m(A) \neq 0$ [14].

Definition 6. A fuzzy measure μ is said to be a p -symmetric measure iff when the coarsest partition of the universal set in subsets of indifference is $\{A_1, \dots, A_p\}$, $A_i \neq \emptyset$, $\forall i \in \{1, \dots, p\}$ [15].

3. Identification of Fuzzy Measures

Several methods of fuzzy measure identification for the Choquet integral have been proposed by solving a specific mathematical programming model. We introduce several popular methods as follows.

3.1. A Maximum Split Approach

Marichal and Roubens [10] proposed a linear programming model to identify fuzzy measures that can be stated as follows:

$$\begin{aligned} & \max \varepsilon \\ \text{s.t.} \left\{ \begin{array}{l} \sum_{T \subseteq S} m_v(T \cup i) \geq 0, \forall i \in N, \forall S \subseteq N \setminus i, \\ t \leq k-1 \\ \sum_{T \subseteq N} m_v(T) = 1, \\ 0 < t \leq k \\ C_{m_v}(u(x)) - C_{m_v}(u(x')) \geq \delta_C + \varepsilon, \\ \text{other constraints.} \end{array} \right. \end{aligned} \quad (7)$$

In this model, m_v is the Möbius representation of a k -additive game μ on N . Note that other constraints include weights on singletons, weights on interactions, and so forth. However, if the above model contains incoherence then the solution set could be empty. A similar situation could result from incomplete information. Hence, we should add extra information to obtain an appropriate solution.

3.2. Minimum Variance Approach

The purpose of the minimum variance method [11] is to maximize the extended Havrda-Charvat entropy of Order 2. The objective function is defined as the variance of the fuzzy measures as:

$$MV(m_v) := \frac{1}{n} \sum_{i \in N} \sum_{S \subseteq N \setminus i} \gamma_s(n) \left(\sum_{T \subseteq S} m_v(T \cup i) - \frac{1}{n} \right)^2 \quad (8)$$

The optimization problem takes the form of the following strictly convex quadratic program:

$$\begin{aligned} & \min MV(m_v) \\ \text{s.t.} \left\{ \begin{array}{l} \sum_{T \subseteq S} m_v(T \cup i) \geq 0, \forall i \in N, \forall S \subseteq N \setminus i, \\ t \leq k-1 \\ \sum_{T \subseteq N} m_v(T) = 1, \\ 0 < t \leq k \\ C_{m_v}(u(x)) - C_{m_v}(u(x')) \geq \delta_C, \\ \text{other constraints.} \end{array} \right. \end{aligned} \quad (9)$$

The above model is a strictly quadratic program and leads to a unique solution because of the convexity of the objective function. However, the model might be useless if the initial preferences are unavailable or give a very uneven Shapley value.

3.3. A Less Constrained Approach

This approach, which derives from the work of Meyer and Roubens [12], can be seen as a generalization of the least squares methods. The minimal preferential information that has to be provided by the decision-maker is a weak order over the available objects. The objective function is defined as:

$$GLS(m_v, y) := \sum_x [C_{m_v}(u(x)) - y(x)]^2 \quad (10)$$

Then, the corresponding mathematical programming model can be described as follows:

$$\begin{aligned} & \min \quad GLS(m_v, y) \\ & \left. \begin{aligned} & \sum_{T \subseteq S} m_v(T \cup i) \geq 0, \forall i \in N, \forall S \subseteq N \setminus i, \\ & t \leq k - 1 \\ & \sum_{T \subseteq N} m_v(T) = 1, \\ & 0 < t \leq k \\ & y(x) - (x') \geq \delta_y, \\ & \text{other constraints.} \end{aligned} \right\} \quad s.t. \end{aligned} \quad (11)$$

Here, δ_y is an indifference threshold and can be considered as the desired minimal difference between the overall utilities of two objects that are viewed as significantly different by the decision-maker. However, the determination of δ_y is a difficult problem since the value of δ_y will affect the solution of the model.

As mentioned previously, although k -additive and p -symmetric measures can reduce the complexity of identifying fuzzy measures, a complicated MCDM problem also makes these methods infeasible. In addition, almost all mathematical programming models need extra information about the preference of the decision-makers to obtain a satisfactory result. In order to apply the Choquet integral in handling a practical regression or classification problem, a data-driven method should be considered.

The most famous algorithm for identifying the fuzzy measures between variables is the heuristic least mean-squares algorithm (HLMS) [1,13,14]. The procedures of the HLMS can be divided into two parts and are described as follows.

Step 0. Initialize the fuzzy capacities at the equilibrium state.

Step 1.1. Given a training datum (x, y) , compute the error function $e = C_{m_v}(u(x)) - y$. Let the values of the fuzzy capacities on the path be involved by x by $u(0), u(1), \dots, u(n)$, where $u(0) = 0$ and $u(n) = 1$.

Step 1.2. Update parameter $u(i)$ to the following equation:

$$u^{new}(i) = u^{old}(i) - \alpha \frac{e}{e_{\max}} (x_{(n-i)} - x_{(n-i-1)})$$

where $\alpha \in (0, 1]$ is the learning rate and $e_{\max}$ is the maximum value of the error. If $y \in [0, 1]$, then $e_{\max} = 1$. The notation $x_{(i)}$ denotes the i th element of the vector x in ascending order.

Step 1.3. For every node $u(i), i = 1, \dots, n-1$, check the monotonicity relations. If $e > 0$, the verification is done for lower neighbors only and if $e < 0$, it is done for upper neighbors only. If a monotonicity relation is violated with $u(k)$, then $u(i) = u(k)$. If a monotonicity relation is violated, such as with node $\mu_j, j \subset X$, then $u(i) = \mu_j$. Repeat steps 1.2 and 1.3 for $i = 1, \dots, n-1$ in the following order:

- If $e > 0$, begin by $u(1), u(2), \dots, u(n-1)$;
- If $e < 0$, begin by $u(n-1), \dots, u(2), u(1)$.

Repeat steps 1.1 to 1.3 for all training data; this is called one iteration. Several iterations can be performed.

Step 2.1. For every node left unmodified in step 1, verify the monotonicity relations with its upper and lower neighbors. If they are not verified, modify the node as in step 1.3.

Step 2.2. For every node left unmodified in step 1, adjust its value considering the values of its upper and lower neighbors in order to have a homogeneous lattice. Let $u(i)$ be the node considered:

- Mean value of upper neighbors $\bar{m}(i) = 1/(n-i) \sum \text{upper neighbors}^{\mu_j}$
- Mean value of lower neighbors $\underline{m}(i) = 1/i \sum \text{lower neighbors}^{\mu_j}$

- Minimum distance between $u(i)$ and its upper (respectively lower) neighbors, denoted as $\bar{d}_{\min}(i)$, (respectively $\underline{d}_{\min}(i)$)

If $\bar{m}(i) + \underline{m}(i) - 2u(i) > 0$, then $u(i)$ is increased:

$$u^{new}(i) = u^{old}(i) + \beta \frac{(\bar{m}(i) + \underline{m}(i) - 2u(i))\bar{d}_{\min}(i)}{2(\bar{m}(i) + \underline{m}(i))} \quad (12)$$

Otherwise $u(i)$ is decreased:

$$u^{new}(i) = u^{old}(i) + \beta \frac{(\bar{m}(i) + \underline{m}(i) - 2u(i))\underline{d}_{\min}(i)}{2(\bar{m}(i) + \underline{m}(i))} \quad (13)$$

and β is a constant value in $(0, 1]$.

Do steps 2.1 and 2.2 for every node left unmodified in the first step. This is called one iteration; several iterations can be performed.

While the HLMS provides a rational and fast procedure to estimate fuzzy measures, it also suffers from the problem of the computational cost incurred with a high-dimensional dataset. Some papers have considered the application of a 2-additive HLMS to deal with this issue [20]. However, a 2-additive HLMS still cannot handle a high-dimensional dataset soundly by reducing the number of fuzzy measures to an acceptable amount. For example, if a dataset contains 50 variables, then the number of the fuzzy measures will be 1275.

4. Hierarchical Choquet Integral

The concept of the hierarchical Choquet integral was proposed by Sugeno and his colleagues [16,21–23] to decompose a Choquet integral model into several sub-Choquet integral models. They gave the definitions of the inter-additive partition (IAP) and inclusion-exclusion covering (IEC) to construct the hierarchical Choquet integral.

Definition 7. A finite measurable partition $\{P_i\}_{i \in \{1, \dots, n\}}$ of X is called an IAP of X if

$$\mu(A) = \sum_{i \in \{1, \dots, n\}} \mu(A \cap P_i) \quad (14)$$

for every $A \in \mathfrak{N}$.

Definition 8. A finite measurable covering $\{C_i\}_{i \in \{1, \dots, n\}}$ of X is called an IEC of X if

$$\mu(A) = \sum_{\substack{I \in \{1, \dots, n\} \\ I \neq \emptyset}} (-1)^{|I|+1} \mu\left(\bigcap_{i \in I} C_i \cap A\right) \quad (15)$$

for every $A \in \mathfrak{N}$.

Suppose that $C = \{C_i\}_{i \in \{1, \dots, n\}}$ is a measurable covering, then each subalgebra S_i of $\mathfrak{N}$ is defined by $S_i := \{C_i \cap A | A \in \mathfrak{N}\}$. Let $M = \{\mu_{C_i}(\cdot)\}_{i \in \{1, \dots, n\}}$ be an n -ary class of non-monotonic fuzzy measures on $S = \{S_i\}_{i \in \{1, \dots, n\}}$. The function f_M on C is defined by

$$f_M(C_i) := (C) \int_{C_i} f d\mu_{C_i} \quad (16)$$

for every measurable function f on $\mathfrak{N}$. Then, the hierarchical Choquet integral model is defined by the model whose input is $f(x)$ on $\mathfrak{N}$, whose output z is represented as $z = (C) \int_C f_M dv$, where v is a non-monotonic fuzzy measure on 2^C .

Then, Sugeno [16] gave the necessary and sufficient condition for an ordinary Choquet integral model and concluded that an overlapping hierarchical Choquet integral model can be hierarchically constructed by an IEC from an ordinary Choquet integral model and a separated hierarchical Choquet integral model can be hierarchically constructed by the IAP from an ordinary Choquet integral model. Hence, information on the IEC and IAP is the key to formulating the hierarchical Choquet integral. Similar to the concept of the IEC and IAP, Tzeng and Huang [24] used the idea of preference separability to construct a hierarchical Choquet integral model.

The concept of the hierarchical Choquet integral is very attractive because it can significantly reduce the number of fuzzy capacities in a model, but the key information is how to divide variables to form sub-Choquet integrals. In addition, the identification of fuzzy measures in real data usually contains errors; however, the hierarchical decomposition theorem above cannot be expected to be satisfied exactly. Furthermore, the identification of IEC, IAP, and preference-separable sets is another difficult problem that arises when dealing with realistic data.

5. GA-Based HLSM

In this paper, we considered our research structure, as shown in Figure 1. First, we considered a real dataset and represented a problem with a hierarchical Choquet integral model. Then, we used Gas to determine the linkage between the input criteria and neurons. Next, we used the HLSM to train the fuzzy measures and calculated the results of the sub-Choquet integral. Furthermore, these results were used to form another regression or classifier to complete a hierarchical Choquet integral for a chromosome. Finally, the processes continued until the stop criterion of GA was met to obtain the final result of the hierarchical Choquet integral.

Figure 1. The research structure of this paper.

In this paper, we considered the hierarchical structure, as shown in Figure 2. In Figure 2, we used genetic algorithms to determine what linkages should be generated from input variables to a sub-Choquet integral. Then, the HLSM was used to derive fuzzy measures and the results of the sub-Choquet integral. Finally, all sub-Choquet integrals were weighted to calculate the final Choquet integral and gave the predicted values.

Figure 2. Hierarchical Choquet Integral.

5.1. GA

The GA was introduced by Holland [25] to mimic the natural evolution of a population by allowing solutions to reproduce chromosomes, create new offspring, and compete for survival in the next generation. In each generation, t , the fittest elements were selected to the mating pool, which was processed by three basic genetic operators for generating new offspring: Reproduction, crossover, and mutation. Based on the principle of the survival of the fittest, we observed that the best chromosome of a sample solution could be obtained. The power of the GA depended on its ability to search for multiple points in a parallel manner instead of a single point. Therefore, the GA could quickly find the appropriate global optimum and avoid reaching a local optimum. The purpose of using the GA in this paper was to determine the input variables required to key into a sub-Choquet integral.

5.2. The GA Procedure

String Representation

To represent the problem, each chromosome was encoded by applying binary strings l_{ij} . The binary value of the ij th genotype (chromosome value) denotes the linkage status from the i th criterion to the j th sub-Choquet integral. Therefore, we should have nm binary strings if we have n criteria and an m sub-Choquet integral.

Population Initialization

The initial population, $P(0)$, could be selected at random between 0 and 1. Each genotype can be initialized to describe the status of the variables from the uniform distribution in the population. Note that there is no standard method for determining the size of the initial population, $P(0)$.

Bhandari et al. [26] showed that, as the number of iterations tends to infinity, the elitist strategy of the GA will provide the optimal string for any size of the initial population, $P(0)$.

Fitness Computation

Here, the objective was to determine the optimal fuzzy measure coefficients by determining, automatically, the HCI with the minimum error between the predicted and the actual values. Hence, the fitness computation could be defined by

$$Accuracy = 1 - \frac{1}{n} \sum_x [HC_{m_v}(u(x)) - y(x)]^2$$

where n denotes the number of instances, $HC_{m_v}(u(x))$ denotes the value of the final HCI value, and $y(x)$ denotes the ground truth class.

5.3. Genetic Operators

Selection. Using the concept of survival of the fittest from natural genetic systems, the selection operator selects chromosomes from the mating pool. Therefore, the good chromosomes get numerous copies, whereas the bad chromosomes die off. The probability of the chromosome's selection is proportional to its fitness value in the population, which is based on the following formula

$$P(x_i) = \frac{f(x_i)}{\sum_{j=1}^N f(x_j)} \quad (17)$$

where $f(x_i)$ denotes the fitness value of the i th chromosome and N denotes the population size. In addition, we used the tournament selection to process the selection operator.

Crossover. The main objective of the crossover was to exchange information between two parent chromosomes and produce two new offspring for the next generation. Several crossover operators could be investigated, such as one-point, two-point, and mask-crossover operators. However, we employed the two-point crossover and set the crossover probability to be equal to 0.8. For example, if we consider three input variables and three sub-Choquet integrals, the design of the crossover here is presented in Figure 3.

The 2nd offspring chromosome can be explained as the first sub-Choquet integral consisting of the input variables x_1 , x_2 , and x_3 . The fuzzy measure coefficients are calculated among these variables.

Mutation. Mutation is a random process in which one genotype is replaced by another to generate a new chromosome. Each genotype has the probability of mutation, P_m , to undergo a mutation change from 0 to 1, and vice versa. We set the mutation probability to be 0.1 in this paper. For example, we can consider a parent and offspring chromosomes with the mutation operator, as shown in Figure 4.

Figure 3. A two-point crossover operator.

Figure 4. The mutation operator.

5.4. Elitist Strategy and Termination Criterion

Elitist Strategy. The elitist strategy carries the fittest chromosome from the previous generation to the next generation. The advantage of the elitist strategy lies in ensuring the selection of the best chromosome and reducing the time of convergence. We set the number of elitists to four in our experiment.

Stop Criterion. The GA typically used two stop criteria: One to set up a maximum number of generations, such as 500 iterations, and the other is triggered when the chromosomes can no longer increase the fitness. We used the first criterion in this paper.

6. Numerical Experiments

In this section, we illustrate the processes of the proposed method and used four datasets to test its effectiveness by considering the regression and classification problems. The first step of the model was to set the parameters of the GA as follows.

6.1. Parameters of GA

For the parameters of the GA in this paper, we employed the following settings, as shown in Table 1.

Table 1. Parameter settings of GA.

Parameters	Value
Gene type	Binary
Population size	50
Number of generations	30
Elitism	4
Crossover probability	0.8
Mutation probability	0.2

In addition, we set two fitness functions for the experiments to consider the tasks of regression and classification, respectively, as follows:

$$fitness\ 1 = -\frac{1}{n} \sum_x [C_{m_v}(u(x)) - y(x)]^2 - \lambda \sum_{j=1}^J (\#(SC_j) - k_j)^2$$

$$fitness\ 2 = -\frac{1}{n} (TP + TN) - \lambda \sum_{j=1}^J (\#(SC_j) - k_j)^2$$

The first term of fitness 1 denotes the mean square error, TP and TN are the numbers of true positive and negative amounts, n is the number of instances, $y(x)$ is the ground truth value, $\#(SC_j)$ denotes the number of input variables in the j th sub-Choquet integral, k denotes the desired number of input variables in the j th sub-Choquet integral, and λ is the penalty parameter.

6.2. Dataset Description

In these experiments, we considered four datasets to demonstrate the proposed method and compared the results with the conventional Choquet integral and classification models—the logit model and neural network. The first two datasets were used for the task of regression and the last two datasets were used for the task of classification. These datasets are described as follows.

The simulated dataset is an artificial dataset that randomly generates positive integers between one and ten. It contains 200 instances and 10 independent variables with one response variable.

The add10 dataset is an artificial dataset gathered from the Delve datasets. It contains 9792 instances and 10 independent variables generated from a uniform distribution between 0 and 1, and one response variable generated by a nonlinear function, which is composed of the first five independent variables. Here, we sampled 500 instances in order to prevent the problem of an expensive computation cost.

The ILPD (Indian liver patient dataset) dataset contains 10 variables that describe the patients' age, gender, total bilirubin, direct bilirubin, and so forth, and the response variable is whether the patients are liver patients or non-liver patients. There are 583 instances comprised of 416 liver patients and 167 non-liver patient records. Here, we only deleted one discrete variable, gender, to retain the remaining nine continuous variables.

The cancer dataset was derived from the University of Wisconsin to record diagnostic breast cancer data. It contains 699 instances and 10 variables, including clump thickness, the uniformity of the cell size, marginal adhesion, and so forth, and all variables are measured with one to ten scores. The response variable is binary (0 for benign and 1 for malignant).

6.3. Experiment Results

We can demonstrate the processes of the proposed method as follows. Take the simulated dataset for example. First, we set the number of the sub-Choquet integral as two and each sub-Choquet contains five input variables. Then, we ran the GA, as described previously, to select the appropriate input variables for each sub-Choquet integral and calculated the weighted Choquet integral to predict the value of the response variable, as shown in Figures 5 and 6.

Figure 5. The convergent process of GA.

Figure 6. Network structure of a hierarchical Choquet integral.

Finally, we could obtain the input variables for each sub-Choquet integral and the weight of the Choquet integral, which is the sum of the two sub-Choquet integrals under the fitness of the mean square error. The results of the Mobius capacity and fitness could be derived, as shown in Table 2.

Table 2. The Mobius capacities and results of the parameters of the example.

2-Additive and 2 Sub-Choquet Integral Each Sub-Choquet Integral Contains 5 Features MSE = 2.359			
Sub-Choquet 1	Mobius Capacity	Sub-Choquet 2	Mobius Capacity
{1}	0.1819	{2}	0.2885
{4}	0.2550	{3}	0.1414
{5}	0.1785	{4}	0.1610
{8}	0.1680	{6}	0.1685
{9}	0.1520	{7}	0.1455
{1,4}	0.1436	{2,3}	−0.0874
{1,5}	0.0268	{2,4}	−0.1140
{1,8}	−0.0039	{2,6}	0.0326
{1,9}	−0.0578	{2,7}	0.0542
{4,5}	−0.1510	{3,4}	0.2016
{4,8}	−0.1012	{3,6}	−0.0534
{4,9}	0.0538	{3,7}	0.0413
{5,8}	0.1337	{4,6}	0.0148
{5,9}	−0.0275	{4,7}	−0.0239
{8,9}	0.0482	{6,7}	0.0292

We could set different hierarchical Choquet integral (HCI) models to calculate the mean square error and compared it with the conventional Choquet integral, as shown in Table 3.

Table 3. Performance comparison between different models and datasets.

Dataset: Simulated Dataset			
Model	# Sub-Choquet Integral	Condition	MSE
2-additive HCI	2	5 variables for each Sub-Choquet Integral	2.359
2-additive HCI	2	6 variables for each Sub-Choquet Integral	2.221
2-additive HCI	2	Unrestricted	2.138
Choquet Integral	0	Unrestricted	2.138
Dataset: Add10 Dataset			
Model	# Sub-Choquet Integral	Condition	MSE
2-additive HCI	2	4 variables for each Sub-Choquet Integral	0.00767
2-additive HCI	2	5 variables for each Sub-Choquet Integral	0.00692
2-additive	2	Unrestricted	0.00663
Choquet Integral	0	Unrestricted	0.00692

Note: # Sub-Choquet Integral denotes the number of sub-Choquet integrals.

It can be seen that the HCI provided a more flexible way to reduce the estimated fuzzy capacities of the variables and outperformance than the conventional Choquet integral used to deal with the task of resolving regression problems.

Next, we could replace our regression model with a classifier to process the task of the classification problem. Here, we considered two popular classifiers, the logit model and multi-layer perceptron (MLP), to serve as the benchmark of the proposed method. Note that we set the MLP model as a three-tier network (i.e., one input layer, one hidden layer with three neurons, and one output layer), and the activation function is the sigmoid function. In addition, we also set three HCI models for each dataset to show the variety of the model's accuracy. The results of the experiment could be presented, as shown in Table 4.

Table 4. The accuracy of the classification models.

Datasets	Logit Model	MLP	2-Additive HCI
cancer	0.9640	0.9710	0.8141 (restrict 4 variables) 0.9531 (unrestricted)
ILPD	0.7074	0.7149	0.7150 (restrict 5 variables) 0.7237 (unrestricted)

From Table 4, it can be seen that, with the release of the number of variables, the accuracy of the model was getting higher. In addition, the proposed method was still competitive with respect to that of the logit model and MLP. Furthermore, the most attractive point of the proposed method is its explain-ability of the model, compared with that of the MLP, which can only be considered as a block box.

7. Discussion

The identification of fuzzy measures played an important role in facilitating the Choquet integral's ability to handle realistic problems. While the Choquet integral has been used for different fields of MCDM, such as feature selection [20], image detection [27], and prediction [28], the huge number of estimated fuzzy measures usually hinders its possible applications. Even though the concept of k -additive, usually 2-additive, fuzzy measures has been proposed to reduce the number of fuzzy measures, it is still a problem when considering really high-dimensional data.

In this paper, we considered a hierarchical Choquet integral to solve the problem of identifying fuzzy measures with high-dimensional data. The main difference between the proposed method and Sugeno and Fujimoto [16] or Tzeng and Huang [24] methods is that, without advance information on the sub-Choquet integrals, the variables will be assigned to the specific sub-Choquet integrals according to the best result of the offspring, with respect to the criterion of MSE or accuracy.

In addition, we compared the proposed method with other conventional approaches by considering the tasks of regression and classification. On the basis of the experimental results, it can be seen that the proposed method is competitive with respect to the other models. In addition, the experimental results also indicated the importance of variable selection in the Choquet integral. That is, not all variables put into the model show the best results.

Finally, the advantages of the proposed method are described as follows. First, the proposed method provides flexibility for a decision-maker to determine how many inputs to add into a sub-Choquet integral. This will be useful when we consider very high-dimensional data. Second, we proposed a data-driven approach to automatically select the most appropriate variables in a sub-Choquet integral. Third, we can simply add a regression or classifier behind the HLMS to consider the task of regression or classification.

8. Conclusions

In this paper, we developed a method to construct the hierarchical Choquet integral by using the GA. The major advantage of the HCI was the reduction in the number of estimated fuzzy capacities between variables. Hence, the proposed method is very suitable for dealing with realistic datasets and enables the Choquet integral to be used for more complicated applications, including the tasks of regression and classification. In addition, the results of the experiments also indicated that the proposed method outperforms the conventional Choquet integral and classifier, namely the logit model and MLP.

Author Contributions: Data curation, C.-Y.C.; Methodology, J.-J.H.; Writing—review & editing, C.-Y.C. and J.-J.H.

Funding: No funding was received.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Grabisch, M. Fuzzy integral in multicriteria decision making. *Fuzzy Set. Syst.* **1995**, *69*, 279–298. [\[CrossRef\]](#)
- Hillier, F. *Evaluation and Decision Models: A Critical Perspective*; Kluwer Academic Publishers: Boston, MA, USA, 2001.
- Sugeno, M. Theory of Fuzzy Integrals and Its Applications. Ph.D. Thesis, Tokyo Institute of Technology, Tokyo, Japan, 1974.
- Zhai, J.; Zang, L.; Zhou, Z. Ensemble dropout extreme learning machine via fuzzy integral for data classification. *Neurocomputing* **2018**, *275*, 1043–1052. [\[CrossRef\]](#)
- Karczmarek, P.; Kiersztyn, A.; Pedrycz, W. On developing Sugeno fuzzy measure densities in problems of face recognition. *Int. J. Mach. Intell. Sens. Signal Process.* **2017**, *2*, 80–96.
- Couceiro, M.; Dubois, D.; Prade, H.; Rico, A. Enhancing the expressive power of Sugeno integrals for qualitative data analysis. In *Advances in Fuzzy Logic and Technology 2017*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 534–547.
- Choquet, G. Theory of capacities. *Ann. de l'institut Fourier* **1954**, *5*, 131–295. [\[CrossRef\]](#)
- Murofushi, T. A technique for reading fuzzy measures (I): The Shapley value with respect to a fuzzy measure. In Proceedings of the 2nd Fuzzy Workshop, Heeze, The Netherlands, 4–5 June 2018; pp. 39–48.
- Murofushi, T.; Soneda, S. Techniques for reading fuzzy measures (III): Interaction index. In Proceedings of the 9th Fuzzy System Symposium, Sapporo, Japan, 24–26 November 1993; pp. 693–696.
- Marichal, J.-L.; Roubens, M. Determination of weights of interacting criteria from a reference set. *Eur. J. Oper. Res.* **2000**, *124*, 641–650. [\[CrossRef\]](#)
- Kojadinovic, I. Minimum variance capacity identification. *Eur. J. Oper. Res.* **2007**, *177*, 498–514. [\[CrossRef\]](#)
- Meyer, P.; Roubens, M. Choice, ranking and sorting in fuzzy multiple criteria decision aid. In *Multiple Criteria Decision Analysis: State of the Art Surveys*; Springer: Berlin/Heidelberg, Germany, 2005; pp. 471–503.
- Grabisch, M. The application of fuzzy integrals in multicriteria decision making. *Eur. J. Oper. Res.* **1996**, *89*, 445–456. [\[CrossRef\]](#)
- Grabisch, M. K-order additive discrete fuzzy measures and their representation. *Fuzzy Sets Syst.* **1997**, *92*, 167–189. [\[CrossRef\]](#)
- Miranda, P.; Grabisch, M.; Gil, P. p-Symmetric fuzzy measures. *Int. J. Unc. Fuzz. Knowl. Based Syst.* **2002**, *10*, 105–123. [\[CrossRef\]](#)
- Sugeno, M.; Fujimoto, K.; Murofushi, T. A hierarchical decomposition of Choquet integral model. *Int. J. Unc. Fuzz. Knowl. Based Syst.* **1995**, *3*, 1–15. [\[CrossRef\]](#)
- Hougaard, J.L.; Keiding, H. Representation of preferences on fuzzy measures by a fuzzy integral. *Math. Soc. Sci.* **1996**, *31*, 1–17. [\[CrossRef\]](#)
- Mesiar, R. Fuzzy measures and integrals. *Fuzzy Set. Syst.* **2005**, *156*, 365–370. [\[CrossRef\]](#)
- Ishii, K.; Sugeno, M. A model of human evaluation process using fuzzy measure. *Int. J. Man-Mach. Stud.* **1985**, *22*, 19–38. [\[CrossRef\]](#)
- Miranda, P.; Grabisch, M. Optimization issues for fuzzy measures. *Int. J. Unc. Fuzz. Knowl. Based Syst.* **1999**, *7*, 545–560. [\[CrossRef\]](#)
- Murofushi, T.; Sugeno, M.; Fujimoto, K. Separated hierarchical decomposition of the Choquet integral. *Int. J. Unc. Fuzz. Knowl. Based Syst.* **1997**, *5*, 563–585. [\[CrossRef\]](#)
- Fujimoto, K.; Murofushi, T.; Sugeno, M. Canonical hierarchical decomposition of Choquet integral over finite set with respect to null additive fuzzy measure. *Int. J. Unc. Fuzz. Knowl. Based Syst.* **1998**, *6*, 345–363. [\[CrossRef\]](#)
- Murofushi, T.; Fujimoto, K.; Sugeno, M. Canonical separated hierarchical decomposition of the Choquet integral over a finite set. *Int. J. Unc. Fuzz. Knowl. Based Syst.* **1998**, *6*, 257–272. [\[CrossRef\]](#)
- Tzeng, G.-H.; Huang, J.-J. *Multiple Attribute Decision Making: Methods and Applications*; Chapman and Hall/CRC: Boca Raton, FL, USA, 2011.
- Holland, J.H. *Adaptation in Natural and Artificial Systems*; University of Michigan: Ann Arbor, MI, USA, 1975; Volume 1, p. 975.
- Bhandari, D.; Murthy, C.; Pal, S.K. Genetic algorithm with elitist model and its convergence. *Int. J. Pattern Recognit. Artif. Intell.* **1996**, *10*, 731–747. [\[CrossRef\]](#)

27. El Baf, F.; Bouwmans, T.; Vachon, B. Fuzzy integral for moving object detection. In Proceedings of the 2008 IEEE International Conference on Fuzzy Systems (IEEE World Congress on Computational Intelligence), Hong Kong, China, 1–6 June 2008; pp. 1729–1736.
28. Yang, R.; Wang, Z.; Heng, P.A.; Leung, K.S. Fuzzified Choquet integral with a fuzzy-valued integrand and its application on temperature prediction. *IEEE Trans. Syst. Man Cybern. Part B (Cybern.)* **2008**, *38*, 367–380. [[CrossRef](#)] [[PubMed](#)]

© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).