mathematics

MDPI

Article

Capacity Random Forest for Correlative Multiple Criteria Decision Pattern Learning

Jian-Zhang Wu *, Feng-Feng Chen, Yan-Qing Li and Li Huang

School of Business, Ningbo University, Ningbo 315211, China; chenfengfengg@163.com (F.-F.C.); hhuliyanqing@163.com (Y.-Q.L.); hlnhym11@163.com (L.H.)

* Correspondence: wujianzhang@nbu.edu.cn; Tel.: +86-150-5883-6418

Received: 17 July 2020; Accepted: 14 August 2020; Published: 16 August 2020

Abstract: The Choquet capacity and integral is an eminent scheme to represent the interaction knowledge among multiple decision criteria and deal with the independent multiple sources preference information. In this paper, we enhance this scheme's decision pattern learning ability by combining it with another powerful machine learning tool, the random forest of decision trees. We first use the capacity fitting method to train the Choquet capacity and integral-based decision trees and then compose them into the capacity random forest (CRF) to better learn and explain the given decision pattern. The CRF algorithms of solving the correlative multiple criteria based ranking and sorting decision problems are both constructed and discussed. Two illustrative examples are given to show the feasibilities of the proposed algorithms. It is shown that on the one hand, CRF method can provide more detailed explanation information and a more reliable collective prediction result than the main existing capacity fitting methods; on the other hand, CRF extends the applicability of the traditional random forest method into solving the multiple criteria ranking and sorting problems with a relatively small pool of decision learning data.

Keywords: fuzzy measure; random forest; ranking and sorting; decision pattern learning; capacity fitting

1. Introduction

The multiple criteria involved in most decision or evaluation problems are usually not independent, and a variety of interactions, from negative to positive, can exist among them [1–3]. The capacity [4], also called fuzzy measure [5], combined with the nonlinear or fuzzy integral [6], especially the most commonly accepted type, the Choquet integral [4], can effectively deal with the complex interaction situations among interdependent criteria and aggregate the multiple-source partial preferences of decision alternatives into their overall assessments, based on which the decision maker can accordingly get the final ranking orders or sorting classifications [7,8].

Given adequate decision instances, generally not a very lot of data, and their overall evaluations, ranking orders, or classifications as the learning set, the scheme of capacity plus Choquet integral can learn, simulate, and explain the correlative multiple criteria decision pattern, wherein the decision and explanation knowledge will be stored in the power set of decision criteria as the capacity values of all decision subsets. This fitting or training process is generally carried out by the linear or nonlinear optimization models [2,9,10], called capacity identification methods [11], among which the least-squares principle method [12–14], the least absolute deviation criterion method [15,16], and the maximum split method [13,17] are the three most widely adopted ones.

However, in the capacity identification method, it is supposed that the decision criteria set is well predetermined by the decision maker and is kept fixed and unchanged during the whole learning and fitting process. Actually, the determination of decision criteria is really a vital precondition of high quality decision making and also a time-consuming task for the decision maker. Furthermore, as we

can imagine, a given unchanged decision criteria set will definitely confine the model's possibility and ability to explore other criteria subsets that enable better performance on learning and explaining the decision pattern behind the training data. Additionally, by using any static decision criteria set, the model does not have the ability to avoid the overfitting problem and also lacks adequately mutual comparison validation with other optional criteria sets. Hence, the traditional capacity identification method can not always ensure the competent generalization ability to the new decision instances, and there is also lack of proof and evidence to verify the reliability and rationality of the prediction result.

The decision tree (DT) [18,19] and random forest (RF) [20,21] framework can help the Choquet capacity and integral decision scheme to improve its abilities of dynamic explanation and reliable prediction. RF is basically an ensemble method of a number of DTs for the purpose of learning and solving classification and regression problems. The outcome of RF is the mean or weighted sum of all selected DTs for the regressive case or their majority vote for the classification case [21]. Different DTs provide different learning and explanation perspectives of the given pattern, and the collective outcome of all DTs with the consideration of their performance will highly increase the credibility and generalization ability of RF's predictions.

RF's classification and regression performance highly depend on the given training data. For the aspect of classification, if with a lot of training data and a small number of desired output categories, RF usually can provide competent performance; but if there is a lack of enough data for each category, like in most situations of the ranking decision problem in which each category only has one instance or record, it is hard for RF to get an acceptable performance. For the aspect of regression, since the interactions exist among decision criteria, the multicriteria decision is basically a nonlinear problem, which is really a challenge for RF to learn and simulate by using some linear regression models. Furthermore, the RF or DT can neither explicitly explain the interaction phenomenon among decision criteria nor specifically express the decision maker's additional judgment and preference for decision criteria and alternatives.

Hence, it is necessary to combine the capacity plus Choquet integral with the DT and RF to better learn and explain the correlative multicriteria decision pattern. With the given decision data and potential decision criteria, q suitable capacity identification method can be used to generate many capacity based DTs (CDTs) and then a capacity-based RF (CRF) can be established by taking account of those trees' performances; see Algorithm 1 for more details.

In summary, there are two direct advantages of CRF. On the one hand, CRF can provide more dynamic explanations and insights of the multicriteria decision pattern, and on the other hand, CRF can have better generalization ability and give a more creditable prediction result using a collective vote of all CDTs. That is, CRF enriches the explanation capability of traditional capacity identification methods, and meanwhile extends the application range of traditional RF into the sorting and ranking multicriteria decision making with a relatively small-scale training dataset.

This paper is organized as follows. After the introduction, we briefly present background knowledge of the capacity, Choquet integral, and capacity identification methods, and of DT and RF in Section 2. In Section 3, we discuss the basic steps of RF and construct the CRF-based decision algorithm and strategies for transforming between ranking and sorting decision problems. In Section 4, we use two illustrative examples to present and analyze the proposed CRF model and algorithm in detail. Finally, we conclude the paper in Section 5.

2. Preliminaries

2.1. Capacity and Its Identification Methods

Let $N = \{1, 2, \ldots, n\}$, $n \geq 2$, be the decision criteria set, $\mathcal{P}(N)$ the power set of N, and $|S|$ the cardinality of subset $S \subseteq N$.

Definition 1. *[4,5,11] A capacity on N is a set function $\mu : \mathcal{P}(N) \to [0,1]$ such that*

(i) $\mu(\emptyset) = 0, \mu(N) = 1$ *(boundary condition);*
(ii) $\forall A, B \subseteq N$, $A \subseteq B$ *implies* $\mu(A) \leq \mu(B)$ *(monotonicity condition).*

The capacity value is generally considered as the importance of criteria subset to the decision problem, and monotonicity means that any new participation of other criteria can not decrease the importance of original coalition [7,11]. Two famous characteristics that stem from the monotonicity with respect to inclusion subsets are the nonadditivity with respect to disjoint subsets and the nonmodularity with respect to any arbitrary two subsets. Nonadditivity and nonmodularity are commonly accepted as the explicit representations of interaction situations among decision criteria [22–24].

The following is one type of probabilistic nonadditivity index (bipartition interaction index), called the Shapley nonadditivity index [23].

Definition 2. *The Shapley nonadditivity index of a subset $A \subseteq N$ is defined as*

$$n^{\mu}_{Sh}(A) = \sum_{B \subseteq N \setminus A} \frac{1}{|N| - |A| + 1} \left(\begin{array}{c} |N| - |A| \\ |B| \end{array} \right)^{-1} \hat{\Delta}_A \mu(B),$$

$$\hat{\Delta}_A \mu(B) = [\mu(B \cup A) - \mu(B)] - \frac{1}{2^{|A|-1}-1} \sum_{\{C, A \setminus C\} \in \pi(A)} ([\mu(B \cup C) - \mu(B)] + [\mu(B \cup (A \setminus C)) - \mu(B)]),$$

$\pi(A) = \{(C, A \setminus C) | C \subset A, C \neq \emptyset, C \neq A\}$ is called the set of proper bipartitions of A.

More properties of nonadditivity indices can be found in [2,22,23,25]. In brief, the Shapley nonadditivity index of a single criterion can be taken as its overall importance index to the decision problem, and the Shapley nonadditivity index of non-singleton nonempty set can be regarded as its comprehensive interaction index, which reflects the interaction kind and tensity among all criteria in it.

A nonlinear or fuzzy integral is a universal notion of the aggregation functions with respect to capacity, among which the Choquet integral is one of the most widely accepted types [4,11].

Definition 3. *For a given $\mathbf{x} \in [0, +\infty]^n$, the discrete Choquet integral of $\mathbf{x}$ with respect to capacity μ on N is defined as:*

$$\mathcal{C}_{\mu}(\mathbf{x}) = \sum_{i=1}^{n} (x_{(i)} - x_{(i-1)}) \mu(\{(i), \ldots, (n)\}),$$

where $x_{(.)}$ is a non-decreasing permutation induced by x_i, $i = 1, \ldots, n$, i.e., $x_{(1)} \leq \ldots \leq x_{(n)}$, and $x_{(0)} = 0$ by convention. The Choquet integral can also be represented in terms of capacity without previously ordering the partial values of $\mathbf{x}$ as [7,26]:

$$\mathcal{C}_{\mu}(\mathbf{x}) = \sum_{A \subseteq N} \hat{x}_A \mu(A).$$

where the basis functions are $\hat{x}_A = \max(0, \min_{i \in A} x_i - \max_{i \in N \setminus A} x_i), \forall A \subseteq N$.

The Choquet integral is an extension of weighted arithmetic mean (WAM) and ordered weight average (OWA) and has some good aggregation properties [6].

The capacity plus Choquet integral can learn and simulate the multiple criteria decision pattern, wherein the decision knowledge is stored as the capacity values of all decision subsets. The knowledge learning and fitting process is usually carried out by some optimization algorithms and models, called the capacity identification methods, which generally take history or typical decision instances and their desired overall evaluations, ranking orders, or sorting classifications as the learning or training dataset. In the following, we introduce three main capacity identification methods.

(1) The least square method

In this method, the decision maker needs to provide the expected overall evaluations of all training alternatives [11]. Suppose the training set as L, and the desired overall evaluation of an alternative $x \in L$ is given as $y(x)$; then the least square method can be constructed as follows:

$$\begin{aligned} &\min z = \sum_{x \in L} (C_\mu(x) - y(x))^2 \\ &\text{the boundary and monotonicity conditions of capacity.} \end{aligned} \tag{1}$$

The objective function is to minimize the square distance between the Choquet integrals with respect to capacity and the given expected overall evaluations of all alternatives in set L. The constraints include the boundary and monotonicity conditions given in the definition of capacity; see Definition 1. It can be figured out that all the constraints are linear but the objective is a quadratic equation—quadratic programming maybe with multiple optimal solutions.

(2) The least absolute deviation method

This method transforms the least square method's nonlinear model into a linear programming by introducing the goal deviation variables [15]. The model is given as:

$$\begin{aligned} &\min z = \sum_{x \in L} d_x^+ + d_x^- \\ &\text{the boundary and monotonicity condition of capacity,} \\ &C_\mu(\mathbf{x}) - d_x^+ + d_x^- = y(x), x \in L, \end{aligned} \tag{2}$$

where L and $y(x)$ are the same as in Equation (1), $d_x^+, d_x^- \geq 0$ are the positive and negative deviation variables and $d_x^+, d_x^- = 0$ implies $C_\mu(\mathbf{x}) = y(x)$. Basically, this model aims to minimize the absolute distance between the obtained Choquet integrals and the desired overall evaluations of all alternatives. This model is linear programming and can be easily solved by most mathematics software, like many linear programming solvers of R packages.

(3) The maximum split method

In this method, the decision maker needs to provide a partial weak order to all training alternatives. Denoted the given preference partial order on L as $O(L)$, the maximum split method can be expressed as follows [27]:

$$\begin{aligned} &\max z = \epsilon \\ &\text{the boundary and monotonicity conditions of capacity,} \\ &C_\mu(x) - C_\mu(x') \geq \epsilon \text{ if } x \succ x' \in O(L). \end{aligned} \tag{3}$$

The above model aims to maximize the distances among all neighbor alternatives in the given order $O(L)$. It obviously uses linear programming and has some advantages in construction and solving. The main drawback is that it contradicts partial order, e.g., $a \succ b$, $b \succ c$, but $c \succ a$ will lead to infeasibility and some inconsistency checking works should be preprocessed; see [8,28,29] for more details of inconsistency recognizing and adjustment.

The above three methods are suitable for solving the alternative ranking decision. As for the sorting decision, we can transform the sorting classification results into the representative overall evaluations of each ordered category or the consistent dominance partial order of all alternatives, and then adopt the above three methods to get the satisfactory capacities; see Section 3.2 for the two strategies.

2.2. The Traditional DT and RF

The DT is a well known machine learning method for solving multiple decision attributes/criteria-based classification and regression problems [18,19] and RF is basically an ensemble methodology of DTs to enhance the prediction correctness rate and the robustness against overfitting [20,21].

With a given learning dataset, the first step is to randomly generate a number of decision trees according to some specific algorithms [30]; then for a new instance, RF assembles all train DTs' outcomes to get a collective prediction, usually a weighted sum for the regressive case or a majority vote for classification case [21]. The collective vote scheme can overcome the low prediction rate of single DT and have a good generalization ability. For a given test dataset, also called out-of-bag (OOB) samples, the generalization error of RF is estimated by OOB error rate in most RF software.

With adequate learning instances, RF can deal with high dimensional classification and regression problems very well and have competent performance compared with other machine learning methods, such as discriminant analysis, support vector machines, and neural networks [20]. However, with little data, the DT and the RF will unavoidably run into the situation of underfitting and cannot provide acceptable and reliable performance.

The multiple criteria decision making problems usually have small learning instance datasets, which are rather inadequate to well train RF and other machine learning methods if part of dataset, usually at least one third of it, needs to be further separated as testing data. However, this kind of decision dataset has two basic characteristics: the first is monotonicity; i.e., if an alternative has larger evaluations on all criteria than another alternative, then its overall evaluation can not be smaller than another one's; the second is nonlinearity; i.e., the overall evaluation of an alternative is generally not a linear representation of its partial evaluations on decision criteria.

As mentioned previously, the capacity identification methods, essentially some optimal models, are competent at fitting this kind of multicriteria decision making pattern. Furthermore, with an importance and interaction index, such as the Shapely nonadditivity index, the relative importance of criteria and interaction situations can be depicted and explained explicitly. Therefore, the capacity plus Choquet integral scheme can help the traditional DT and RF to well deal with the challenges from the monotonicity and nonlinearity lying in the small scale of the decision dataset.

3. The CRF Decision Method

In this section, we combine the random forest framework with the capacity plus Choquet integral scheme to establish the capacity random forest method to better solve and explain the decision ranking and sorting problems.

3.1. The CRF Algorithm for the Ranking Decision Problem

The ranking decision dataset generally includes the partial evaluations of alternatives on all criteria and their overall evaluations or their ranking orders. As mentioned before, the least square method, Model (1), and least absolute deviation method, Model (2), are competent for learning the overall evaluation type decision data, and maximum split method is suitable for the ranking order-type decision data. Combining these capacity identification methods with the RF framework, we can have the CRF algorithm; see Algorithm 1.

In CRF Algorithm 1, the condition $|S| \leq k$, k can be empirically set as 6 on account of the exponential complexity inherent in the construction of capacity; more precisely, there are $2^n - 2$ coefficients involved in the capacity and Choquet integral for n decision criteria. When $n \leq 6$, we can just set $k = n$ as well.

Algorithm 1 Capacity random forest (CRF) algorithm for ranking decisions.

Input: Decision criteria set N, learning set L, the tree number t of the forest, the largest number k of criteria in a tree.

Output: The capacity random forest and the prediction of new instance.

```
/* Training of the capacity random forest.  */
```

for S *in* $\mathcal{P}(N)$ **do**
  if $|S| \geq 2$ *and* $|S| \leq k$ **then**
    Use capacity identification method to get the optimal capacity on S, denoted as μ_S.
    Calculate the performance of μ_S, denoted as $f(\mu_S)$.
  end
  Based on performances $f(\mu_S)$, identify the appearance frequency of each tree $q(\mu_S)$ and randomly generate t trees to get the capacity random forest.
end

```
/* Predicting by the capacity random forest.  */
```

for *every new unknow instance* **do**
  Calculate the outcomes of all the decision trees of given new instance.
  Aggregate these outcome into the collective prediction evaluation.
end

Get the final ranking orders of all new instances according to their collective predictions.

In CRF Algorithm 1, the performance of each CDT about S, or simply of the optimal capacity μ_S, should deeply depend on the objective function value of the adopted capacity identification method, which is denoted as z_μ. Since the Models (1) and (2) are able to minimize their objective functions, we can set their CDT's performance function as:

$$f(\mu_S) = \begin{cases} 1/z_\mu & \text{if } z_\mu \neq 0 \\ 1/z_0 & \text{if } z_\mu = 0 \end{cases} \tag{4}$$

where $z_0 = \min(\eta_1, \frac{\min_{s \in S, z_\mu \neq 0} z_\mu}{\eta_2})$ is the adjusted objective function value of CDT whose model's optimal objective function is of zero; η_1 is the expected best objective function value of all CDTs; η_2 is the least ratio between the minimum of nonzero objective function values of all CDTs and best objective function value. Since Model 3 aims to maximize, we can set the performance function of CDT with the positive objective function value as

$$f(\mu_S) = z_\mu \tag{5}$$

Then the appearance frequency of each CDT in the CRF is defined as:

$$q(\mu) = f(\mu_S) / \sum_{2 \leq |A| \leq k} f(\mu_A). \tag{6}$$

Remark 1. *In Equation (4), z_0 is adopted to reasonably confine the appearance frequencies of CDTs with zero objective function, which is connected to the diversity of CDTs and can help to maintain the good generalization ability of CRF and avoid the case of overfitting. Empirically, we can set $\eta_1 = 10^{-4}$ and $\eta_2 = 10^2$. For example, if $\min_{s \in S, z_\mu \neq 0} z_\mu > 10^{-2}$, e.g., 0.02, then $z_0 = \min(10^{-4}, 0.02/100) = 10^{-4}$; if $\min_{s \in S, z_\mu \neq 0} z_\mu < 10^{-2}$, e.g., 0.002, then $z_0 = \min(10^{-4}, 0.002/100) = 2 \times 10^{-5}$. The Equation (5) only involves the CDTs whose objective functions are positive, mainly because zero or negative objective function values mean Model (3) fails on splitting the decision alternatives and then we just abandon those incompetent CDTs.*

In CRF Algorithm 1, the number of decision trees, t, should be lager than 2^k at least and in general with a scale of hundreds. The prediction of overall evaluation and rank can adopt the simple arithmetic

average of the outcomes of those trees, because the training process of the capacity random forest has already considered those trees' performances through their appearance frequencies.

3.2. The CRF Algorithm for the Sorting Decision Problem

The sorting decision problem aims to classify all decision alternatives into a set of ordered classes. Since a lack of training instances likely causes the traditional RF an underfitting case, we need to transform the sorting decision into the overall evaluation or ranking order-types of decision problem.

Mathematically, supposing there are m sorted classes $C = \{C_1, ..., C_m\}$, where if $x \in C_j, x' \in C_{j+1}$, $j = 1, ..., m-1$, then $x \succ x'$, we can have the following two strategies:

(a) Construct a series of partial orders: $C_\mu(x) \geq C_\mu(x') + \delta, \delta > 0, \forall x \in C_j, \forall x' \in C_{j+1}, j = 1, ..., m-1$, and then adopt model (3) to solve this transformed ranking decision problem.

Or alternatively,

(b) Set the representative overall evaluations of m classes as $e_1, ..., e_m$, where $e_j \geq e_{j+1} + \delta$, $\delta > 0$, $j = 1, ..., m-1$, and adopt model (1) or (2) to identify the desired capacity;

With the above two transformation strategies, the CRF Algorithm 1 can be applied smoothly to solve the sorting decision problem.

Strategy (a) is to transform the sorting decision problem into the ranking-order decision. In strategy (a), it is necessary to define all dominant relationships between the decision alternatives in all neighbor classes, and the threshold δ is a small positive number, e.g., 0.05 to differentiate the ordered classes. One can figure out, in strategy (a), the total number of partial order constraints depends on the number of classes and the number of elements in each class; more specially, the total number of constraints is $\sum_{i=1}^{m-1} |C_i| \times |C_{i+1}|$. Some reduction methods for these types of constraints can be found in [13,27]. Another potential issue for strategy (a) is that some inconsistencies may possibly exist in these hard constraints, and in this case, the inconsistency check and adjustment methods need be applied before identifying the desired capacity; more technologies about inconsistency check and adjustment strategies can be found in [25,28,31].

Strategy (b) aims to transform the sorting decision problem into the overall evaluation decision, where δ acts the same role as in strategy (a). In strategy (b), the critical task is to find the suitable representative overall evaluation of each class, which needs the decision maker to have some background knowledge about the real decision problem and also a few rounds of trial and error. Fortunately, this task has much good freedom, and it is not difficult to obtain an acceptable result because of the inherent property of multiple criteria decision making and the characteristics of capacity identification models. On the one hand, the monotonicity or nondecreasing property of the aggregation function in decision context can ensure that the positive threshold between the desired representative evaluations of different classes can have relatively large allowable ranges; see the illustrative example in Section 4.3. On the other hand, in the corresponding capacity identification models, see Models (1) and (2), these representative values are only involved in objective function or soft goal constraints; as a result, the rigorous inconsistency checking can be omitted accordingly. Meanwhile, the number of constraints should be rather less than that of strategy (a); see the analysis in Section 4.3.

4. Two Illustrative Examples

4.1. Students Evaluation Example

In this subsection, we adapt the example in [11]. The decision maker uses five subjects (criteria) (1) statistics, (2) probability, (3) economics, (4) management, and (5) English, to evaluate seven students, $a \sim g$, whose scores with scale $[0, 20]$ are given in Table 1.

Table 1. Scores of seven students on five subjects.

	1	**2**	**3**	**4**	**5**
a	18	11	11	11	18
b	18	11	18	11	11
c	11	11	18	11	18
d	18	18	11	11	11
e	11	11	18	18	11
f	11	11	18	11	11
g	11	11	11	11	18

The decision maker sets the desired overall evaluations of seven students as:

$$15.0, 14.5, 14.0, 13.5, 13, 12.5 \text{ and } 12. \tag{7}$$

With these partial and desired overall evaluations of seven students as the training set, we ran the Algorithm 1 by adopting the least absolute deviation method, i.e., Model (2), as the capacity identification method. Table 2 shows the results of $k = 2, 3, 4, 5$ and $t = 20, 200, 2000, 20,000$, wherein first column holds the values of k, the second column has values of t, the third column is the running time (seconds) of the algorithm on a common laptop with CPU i3-4030U 1.90 GHz and RAM 4.00 GB, the fourth to tenth columns are the prediction values of seven students by the corresponding capacity random forests, and the last column titled by "Obj" shows the sums of absolute deviations, i.e., absolute distances, between the predicted evaluates and the desired evaluations. According to the last column, one can see the 200-tree forests seem to be good enough for the random forest with any k for this example. When $k = 4$ and 5, the "Obj" values are all 0; the main reason for this fact is that by criteria subsets $\{1, 3, 4, 5\}$ and $\{1, 2, 3, 4, 5\}$, their corresponding models' objective function values are 0, so according to Equation (4), these two subsets CDTs' appearance frequencies in CRF will be overwhelmingly superior to other CDTs and the final outcome of CRF will be mainly determined by these two superior CDTs.

Table 2. Results of Algorithm 1 with different parameters.

k	***t***	**Time (s)**	**a**	**b**	**c**	**d**	**e**	**f**	**g**	**Obj**
2	20	1.80	13.7250	13.6000	13.0500	13.5250	12.4000	11.6000	12.0500	4.7000
2	**200**	5.86	13.4275	13.7300	13.1450	13.3175	12.8650	12.0050	11.7400	**4.2700**
2	2000	6.51	13.4970	13.6638	13.0170	13.3422	12.8260	11.9200	11.7410	4.4930
2	20,000	463.57	13.4921	13.5623	13.0071	13.3679	12.7774	11.8858	11.7653	4.6420
3	20	6.86	14.0000	14.1250	12.9750	13.5250	12.5500	12.2000	11.6000	3.5750
3	**200**	7.23	14.2725	14.0750	13.3425	13.5100	12.5725	12.2175	11.8200	**2.7100**
3	2000	11.81	14.1595	13.9298	13.3665	13.4495	12.6337	12.2050	11.8560	2.9000
3	20,000	475.47	14.1277	13.9580	13.3616	13.4283	12.6706	12.2198	11.8388	2.8953
4	20	14.94	15.0000	14.5000	14.0000	13.5000	13.0000	12.5000	12.0000	0
4	**200**	15.16	15.0000	14.5000	14.0000	13.5000	13.0000	12.5000	12.0000	**0**
4	2000	19.30	15.0000	14.5000	14.0000	13.5000	13.0000	12.5000	12.0000	0
4	20,000	491.23	15.0000	14.5000	14.0000	13.5000	13.0000	12.5000	12.0000	0
5	20	19.17	15.0000	14.5000	14.0000	13.5000	13.0000	12.5000	12.0000	0
5	**200**	20.38	15.0000	14.5000	14.0000	13.5000	13.0000	12.5000	12.0000	**0**
5	2000	24.68	15.0000	14.5000	14.0000	13.5000	13.0000	12.5000	12.0000	0
5	20,000	507.60	15.0000	14.5000	14.0000	13.5000	13.0000	12.5000	12.0000	0

Remark 2. *For the purpose of comparison, we ran the traditional RF algorithm in R with package "randomForest" version 4.6-14. If we adopt Equation (7) as the training target values, the random forest automatically uses regression method to fit this decision problem. With 500 decision trees and other parameters as the default, the RF package gives the fitted results of seven students as:*

$$14.06568, 13.99443, 13.58968, 13.64960, 13.03828, 13.03450 \textit{ and } 13.07630.$$

One can figure out that the above result was not successful at simulating the decision pattern, especially since the student g has a slight dominance over students e and f, which is somewhat unacceptable. Compared to the result in Table 2, one can conclude that the CRF is more competent than the traditional RF for fitting this decision example.

Now, we consider the partial order case decision. Suppose that the decision maker provides a partial weak order of the seven students as:

$$a \succ b \succ c \succ d \succ e \succ f \succ g, \tag{8}$$

which is consistent with the previously given desired overall evaluations of seven students and the partial order generated by last eight rows in Table 2. By taking data in Table 1 and simultaneously adopting the maximum split method, i.e., Model (3), as the capacity identification method, we can carry out the CRF method; Algorithm 1. It can be verified that only the CDTs of criteria subsets $\{1,3,4,5\}$ and $\{1,2,3,4,5\}$ can correctly differentiate the seven students with the same threshold 1.166667 (when adopting other criteria subsets, the optimal objection function values of Model (3) are all 0 or negative) and output the same predictions of students $a \sim g$ as:

$$18.00000, 16.83333, 15.66667, 14.50000, 13.33333, 12.16667 \text{ and } 11.00000, \tag{9}$$

which is also the same ranking orders by Algorithm 1 with $k = 4$ and 5.

Furthermore, we can get more explanations and decision information from the CRF algorithm. For example, the decision subset $\{1,3,4,5\}$ is basically good enough to fit the above two kinds of ranking decision problems, which is very useful if the decision maker wants to carry out a reduction on the decision criteria. The optimal capacity on $\{1,3,4,5\}$ and its Shapely nonadditivity index are listed in Table 3. One can see that criterion 1 has the largest importance, 0.3869; criteria 3 and 5 have similar importance (0.2560 and 0.2440); and criteria 4 has the least importance, 0.1131, in this decision process of student evaluation. As for the interaction among criteria, almost all criteria have positive interactions (see, e.g., the nonadditivity of these four criteria, 0.2755), but criteria {1, 3} have a slightly negative interaction; see its nonadditivity index of –0.0476.

Table 3. The capacity on {1,3,4,5} and its Shapley nonadditivity index.

A	$\mu(A)$	$n^{\mu}_{\mathrm{Sh}}(A)$
$\emptyset$	0.0000	0.4155
{1}	0.3571	0.3869
{3}	0.2143	0.2560
{4}	0.0000	0.1131
{5}	0.1429	0.2440
{1, 3}	0.5000	–0.0476
{1, 4}	0.3571	0.1310
{3, 4}	0.2857	0.1667
{1, 5}	0.5714	0.1310
{3, 5}	0.4286	0.0952
{4, 5}	0.1429	0.1310
{1, 3, 4}	0.5000	0.0833
{1, 3, 5}	0.5714	0.0357
{1, 4, 5}	0.5714	0.1786
{3, 4, 5}	0.4286	0.1786
{1, 3, 4, 5}	1.0000	0.2755

Actually, according to the partial order (8), we can divide the seven students into seven classes such that each class has only one student, and in this sense, the ranking decision essentially turns into

a special sorting decision. See the following quality evaluation problem in the next subsection; it is a more general case and each class has more than one decision alternative.

Remark 3. *For the purpose of comparison, we also ran the traditional RF algorithm with R package "randomForest" version 4.6-14 by taking each student as a different class; i.e., take the learning result as a "factor" type datum with seven elements. Unfortunately, the results of the traditional random forest are very messy; the OOB (out of bag) estimate of the error rate was 100% with 500 decision trees and default parameter values of classification. This means the traditional RF fails in fitting this decision problem and the main reason should have been the extreme lack of training data in each class.*

4.2. Quality Evaluation Example

This example was adapted from [27]. The decision maker plans to sort 20 alternatives (see Table 4) into three classes (good, medium and bad) according to eight criteria, and also gives some reference information (i.e., the training dataset) as follows:

- CenConf, Loe, OM, EII belong to class good;
- Zhang, ECR, PDI belong to class medium;
- Sup, Surp, Kappa belong to class bad.

Table 4. The partial evaluations of 20 alternatives on 8 criteria.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Sup	0	0	0	0	1	0	1	2
Conf	1	0	0	1	1	0	1	2
R	0	1	1	0	1	0	1	1
CenConf	1	1	1	0	1	0	1	2
PS	0	1	1	0	1	1	1	1
Loe	1	1	1	1	1	0	1	1
Zhang	1	1	1	1	2	0	0	0
ImpInd	1	1	1	0	1	1	1	0
Lift	0	1	1	0	1	0	1	1
Surp	1	1	0	0	1	0	1	1
Seb	1	0	0	1	0	0	1	1
OM	1	1	1	1	0	0	1	2
Conv	1	1	1	1	0	0	1	1
ECR	1	0	0	1	2	0	1	1
Kappa	0	1	1	0	1	0	1	0
IG	0	1	1	0	2	0	1	0
IntImp	1	1	1	1	2	1	1	0
EII	1	1	1	1	2	1	0	0
PDI	1	1	1	0	1	1	1	0
Lap	1	0	0	0	1	0	1	0

Hence, we can adopt the strategy (a) given in Section 3.2 to transform the above references information into partial order constraints as the alternatives in class good dominate the alternatives in class medium and those in class medium dominate those in class Bad:

$$\text{CenConf} \succ \text{Zhang}, \text{CenConf} \succ \text{ECR}, \text{CenConf} \succ \text{PDI}, \cdots, \text{PDI} \succ \text{Kappa}. \tag{10}$$

Additionally, while taking the above dominant constraints and adopting the maximum split model, Model (3), we ran CRF Algorithm 1 and found out that the subsets in Table 5 have non zero splits, which means with these subsets' CTDs, the capacity plus Choquet integral patterns, can correctly differentiate the given alternatives. For example, CDT of {3, 4, 6, 8} can split all these learning stances very well with the largest threshold 0.5; the importance and interaction indices of the criteria subsets are listed in Table 6.

Table 5. The decision subsets with splits larger than zero.

Subsets	$\{3,4,6,8\}$	$\{3,5,6,8\}$	$\{1,3,4,6,8\}$	$\{2,3,4,6,8\}$
Splits	0.5	0.3333	0.5	0.5
Subsets	$\{1,3,5,6,8\}$	$\{2,3,5,6,8\}$	$\{2,4,5,6,8\}$	$\{3,4,5,6,8\}$
Splits	0.3333	0.3333	0.1667	0.5
Subsets	$\{3,4,6,7,8\}$	$\{3,5,6,7,8\}$	$\{1,2,3,4,6,8\}$	$\{1,2,3,5,6,8\}$
Splits	0.5	0.3333	0.5	0.3333
Subsets	$\{1,2,4,5,6,8\}$	$\{1,3,4,5,6,8\}$	$\{2,3,4,5,6,8\}$	$\{1,3,4,6,7,8\}$
Splits	0.1667	0.5	0.5	0.5
Subsets	$\{2,3,4,6,7,8\}$	$\{1,3,5,6,7,8\}$	$\{2,3,5,6,7,8\}$	$\{2,4,5,6,7,8\}$
Splits	0.5	0.3333	0.3333	0.1667
Subsets	$\{3,4,5,6,7,8\}$	$\{1,2,3,4,5,6,8\}$	$\{1,2,3,4,6,7,8\}$	$\{1,2,3,5,6,7,8\}$
Splits	0.5	0.5	0.5	0.3333
Subsets	$\{1,2,4,5,6,7,8\}$	$\{1,3,4,5,6,7,8\}$	$\{2,3,4,5,6,7,8\}$	$\{1,2,3,4,5,6,7,8\}$
Splits	0.1667	0.5	0.5	0.5

Table 6. The capacity on {3,4,6,8} and its Shapley nonadditivity index.

A	$\mu(A)$	$n^{\mu}_{\mathrm{Sh}}(A)$
$\emptyset$	0.0000	0.0000
{3}	0.0000	0.5000
{4}	0.0000	0.1667
{6}	0.0000	0.0833
{8}	0.0000	0.2500
{3, 4}	0.5000	–0.0000
{3, 6}	0.5000	0.2500
{4, 6}	0.0000	0.0000
{3, 8}	1.0000	0.2500
{4, 8}	0.5000	0.0000
{6, 8}	0.0000	–0.2500
{3, 4, 6}	1.0000	0.1667
{3, 4, 8}	1.0000	0.1667
{3, 6, 8}	1.0000	0.1667
{4, 6, 8}	0.5000	–0.1667
{3, 4, 6, 8}	1.0000	0.1429

The prediction results of 20 alternatives of CRF with different levels k from 4 to 8 and $t = 200$ are given in Table 7. According to the results of $k = 8$, we give the following classification rules: the alternative is good if its prediction value is larger than 0.9; medium if less than 0.9 and larger than 0.45; bad if less than 0.45; see the last second column "CRF", which stands for "capacity random forest", in Table 7 for their predicted classes.

Remark 4. *For the purpose of comparison, we ran the traditional random forest by R with "randomForest" version 4.6-14 for the quality evaluation example. With 500 DTs and other parameters as defaults, the RF's OOB estimate of classification error rate was 50%. The prediction results of all alternatives are given in last column "TRF", which stands for "traditional random forest", in Table 7, which are similar to the results of CRF; see the second to last column in Table 7, and italics in last column show two different results between CRF and TRF; see third and ninth rows. Besides, we should point out that the running times of CRF, Algorithm 1, with $k = 4, 5, 6, 7, 8$ in Table 7, were 235.45, 703.91, 1512.15, 2346.31, and 2716.71 s, respectively (still by the laptop with CPU i3-4030U 1.90 GHz and RAM 4.00 GB). That means the efficiency of CRF is not comparable with that of traditional RF. The main reason for such a time consuming situation is the exponential complexity associated with capacity, or equally the construction process of each CDT.*

Table 7. The prediction results with $k = 4 \sim 8$ and $t = 200$.

k	4	5	6	7	8	CRF	TRF
Sup	0.1500	0.1000	0.1617	0.1550	0.0950	**Bad**	**Bad**
Conf	0.4250	0.4500	0.4842	0.5092	0.4958	Medium	Medium
R	1.0000	0.9167	0.6517	0.5633	0.5767	Medium	*Bad*
CenConf	1.0000	1.0000	1.0000	1.0000	1.0000	**Good**	**Good**
PS	1.0000	0.9167	0.6608	0.5692	0.5808	Medium	Medium
Loe	1.0000	1.0000	1.0000	1.0000	1.0000	**Good**	**Good**
Zhang	0.5750	0.5500	0.5808	0.5775	0.5475	**Medium**	**Medium**
ImpInd	0.5750	0.5500	0.5808	0.5775	0.5475	Medium	Medium
Lift	1.0000	0.9167	0.6517	0.5633	0.5767	Medium	*Bad*
Surp	0.1500	0.1000	0.1617	0.1550	0.0950	**Bad**	**Bad**
Seb	0.2750	0.3250	0.2192	0.2092	0.1942	Bad	Bad
OM	1.0000	1.0000	1.0000	1.0000	1.0000	**Good**	**Good**
Conv	1.0000	1.0000	0.9733	0.9683	0.9833	Good	Good
ECR	0.5750	0.5500	0.5808	0.5775	0.5475	**Medium**	**Medium**
Kappa	0.1500	0.1000	0.1617	0.1550	0.0950	**Bad**	**Bad**
IG	0.3000	0.2000	0.2850	0.2550	0.1633	Bad	Bad
IntImp	1.0000	1.0000	1.0058	1.0025	1.0025	Good	Good
EII	1.0000	1.0000	1.0000	1.0000	1.0000	**Good**	**Good**
PDI	0.5750	0.5500	0.5808	0.5775	0.5475	**Medium**	**Medium**
Lap	0.1500	0.1000	0.1233	0.1000	0.0683	Bad	Bad

4.3. Further Discussions on Targeted Evaluations of Ordered Classes for Sorting Decisions

Previously, we have applied strategy (a) mentioned in Section 3.2 to transform the learning datasets of two decision examples into ranking order-types of data; see Equations (8) and (10). Next, we further discuss the application of the strategy (b) and mainly investigate how to set the desired evaluations of the ordered classes.

A basic fact or axiom in multiple criteria decision analysis is that if alternative x is no less than alternative y on every criterion, then the overall evaluation (e.g., the Choquet integral value) of x should be no less than that of y, which is also called the nondecreasing or monotonicity property of aggregation function. From this point of view, we can empirically select a series of different numbers with a certain threshold from the common range of alternatives, and then sort them in an ascending/descending order as the targeted evaluations of the ordered classes.

For example, we can take the target values in Equation (7) and those in Equation (9) as the values of seven students desiring to run the CRF algorithm with the least square method or least absolute deviation method. Additionally, it can be verified that these target values have relatively large freedoms: if we divide the seven students into three groups as $\{a, b\}, \{c, d, e\}$ and $\{f, g\}$ and take $\{18, 14, 11\}$ as their targeted evaluations, the capacity of the random forest can fit this ranking problem with zero deviation, so do well with targeted evaluation of $\{16, 14, 12\}$ or even $\{18, 17, 16\}$.

Furthermore, for the quality evaluation of 20 alternatives, we give three cases with $\{1, 0.5, 0\}$, $\{2, 1, 0\}$ and $\{0.5, 0.3, 0.1\}$ as their targeted evaluations; the corresponding results by the capacity random forest algorithm are given in Table 8, wherein first three rows of the "Intervals" columns are the allowable values that belong to the three classes; e.g., the element of first row and third column, [0.6, 1], means if the prediction value falls in this interval, the alternative will belong to "good." Additionally, these intervals are also given empirically. The italic words in Table 8 mean the infliction with result of the last column in Table 7.

One can figure out that the strategy (b) has a rather equivalent performance to strategy (a), which greatly comes from the flexibility and nonlinearity of capacity and Choquet integral decision pattern. Finally, we exemplify that strategy (b) has some advantages over strategy (a) in this application:

- It needs relatively fewer constraints, e.g., 7 constraints of strategy (b) comparing with the $4 \times 3 + 3 \times 3 = 21$ constraints of strategy (a) for the quality evaluation problem;

- Strategy (b) basically uses the goal constraints which are soft equations and does not further concern the inconsistency among these constraints or the infeasibility issue in application; however, the strategy (a) has to face the strict constraints and carefully deals with the unavoidable contradiction among them (some inconsistency checks and adjustment strategies can be found in references [8,25,32]).

Table 8. The prediction results with different target values.

	Aimed Values	**Intervals**	**Aimed Values**	**Intervals**	**Aimed Values**	**Intervals**
Good	1	[0.6, 1]	2	[0.9, 1.2]	0.5	[0.35, 0.7]
Medium	0.5	[0.3, 0.6)	1	[0.3, 0.9)	0.3	[0.2, 0.3)
Bad	0	[0, 0.3)	0	[0, 0.3)	0.1	[0, 0.2)
Sup	0.0000	Bad	0.3875	*Medium*	0.1000	Bad
Conf	0.4775	Medium	0.9475	*Good*	0.2555	Medium
R	0.3050	Medium	0.3775	Medium	0.2140	Medium
CenConf	1.0000	Good	0.9500	Good	0.6075	Good
PS	0.3050	Medium	0.4850	Medium	0.2140	Medium
Loe	1.0000	Good	0.9925	Good	0.6875	Good
Zhang	0.5000	Medium	0.9925	*Good*	0.3000	Medium
ImpInd	0.5000	Medium	0.9075	*Good*	0.3000	Medium
Lift	0.3050	Medium	0.3775	Medium	0.2140	Medium
Surp	0.3850	*Medium*	0.3450	*Medium*	0.2360	*Medium*
Seb	0.3200	*Medium*	0.5625	*Medium*	0.1885	Bad
OM	1.0000	Good	1.1625	Good	0.6875	Good
Conv	1.0000	Good	0.9875	Good	0.6875	Good
ECR	0.4775	Medium	0.8525	Medium	0.2890	Medium
Kappa	0.0000	Bad	0.1925	Bad	0.1000	Bad
IG	0.0000	Bad	0.2725	Bad	0.1335	Bad
IntImp	0.6525	Good	1.0575	Good	0.3750	Good
EII	0.6525	Good	1.0500	Good	0.3750	Good
PDI	0.5000	Medium	0.9075	*Good*	0.3000	Medium
Lap	0.0875	Bad	0.1875	Bad	0.0815	Bad

5. Conclusions

In this paper, we combined the traditional RF framework with the capacity and Choquet integral decision model to develop the CRF scheme, in which the capacity plus Choquet integral is taken as CDT while using the capacity values to store the decision knowledge and the capacity identification or fitting methods as the suitable knowledge learning and model training tools. With the CRF algorithm, more decision aids and explanation information, such as the fitting abilities of decision criteria subsets, the proper criteria reduction suggestion, and the importance and interactions situation among specific criteria subsets, can be obtained accordingly. Meanwhile, CRF has good generalization ability and its final prediction result becomes relatively persuadable and confidential as a collective vote outcome. It was proven that, compared to traditional RF, the CRF is competent to deal with the small scale of a decision learning dataset because of the flexibility and nonlinearity inherent in the capacity and Choquet integral scheme.

It should be admitted that if with an adequate or large amount of decision learning data, especially for a sorting decision problem with a small number of classifications, the traditional DT and RF can have much better performance than CDT and CRF. The exponential complexity inherent in the structure of capacity will limit its efficiency in dealing with a large scale dataset. Hence, in the future, we plan to adopt more special families of capacities, e.g., the k-additive capacity [33], the k-maxitive and minitive capacity [26,34], the k-interactive capacity [35], and the k-order representative capacity [36], to reduce the construction complexity, and also we will enrich the types of CDTs and CRFs to enhance their fitting ability and flexibility by adopting more forms of nonlinear or fuzzy integrals, such as the Sugeno integral [5], the pan integral [37], and the inclusion-exclusion integral [38].

Author Contributions: Methodology, J.-Z.W.; Software, J.-Z.W., L.H. and F.-F.C.; Writing—original draft, J.-Z.W. and Y.-Q.L.; writing—review and editing, J.-Z.W., L.H. and Y.-Q.L. All authors have read and agreed to the published version of the manuscript.

Funding: The work was supported by the National Natural Science Foundation of China (number 71671096).

Acknowledgments: The authors are grateful to the anonymous reviewers for their valuable comments and suggestions which have led to significant improvements in this paper.

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

References

1. Ehrgott, M.; Figueira, J.; Greco, S. *Trends in Multiple Criteria Decision Analysis*; Springer: Berlin/Heidelberg, Germany, 2010.
2. Beliakov, G.; James, S.; Wu, J.Z. *Discrete Fuzzy Measures: Computational Aspects*; Springer: Cham, Switzerland, 2019.
3. Grabisch, M. Fuzzy integral in multicriteria decision making. *Fuzzy Sets Syst.* **1995**, *69*, 279–298. [CrossRef]
4. Choquet, G. Theory of capacities. *Annales de l'institut Fourier* **1954**, *5*, 131–295. [CrossRef]
5. Sugeno, M. Theory of Fuzzy Integrals and Its Applications. Ph.D. Thesis, Tokyo Institute of Technology, Tokyo, Japan, 1974.
6. Grabisch, M. *Set Functions, Games and Capacities in Decision Making*; Springer: Berlin, Germany; New York, NY, USA, 2016.
7. Beliakov, G.; Bustince Sola, H.; Calvo, T. *A Practical Guide to Averaging Functions*; Springer: New York, NY, USA, 2016.
8. Wu, J.Z.; Xi, R.J.; Beliakov, G. Nonadditivity Index Oriented Decision Preference Information Representation and Capacity Identication. *Econ. Comput. Econ. Cybern. Stud. Res.* **2020**, *54*, 281–297.
9. Purcaru, C.; Precup, R.E.; Iercan, D.; Fedorovici, L.O.; David, R.C.; Dragan, F. Optimal robot path planning using gravitational search algorithm. *Int. J. Artif. Intell.* **2013**, *10*, 1–20.
10. Goli, A.; Aazami, A.; Jabbarzadeh, A. Accelerated cuckoo optimization algorithm for capacitated vehicle routing problem in competitive conditions. *Int. J. Artif. Intell.* **2018**, *16*, 88–112.
11. Grabisch, M.; Kojadinovic, I.; Meyer, P. A review of methods for capacity identification in Choquet integral based multi-attribute utility theory: Applications of the Kappalab R package. *Eur. J. Oper. Res.* **2008**, *186*, 766–785. [CrossRef]
12. Grabisch, M.; Labreuche, C. A decade of application of the Choquet and Sugeno integrals in multi-criteria decision aid. *Ann. Oper. Res.* **2010**, *175*, 247–286. [CrossRef]
13. Meyer, P.; Roubens, M. Choice, Ranking and Sorting in Fuzzy Multiple Criteria Decision Aid. In *Multiple Criteria Decision Analysis: State of the Art Surveys*; Figueira, J., Greco, S., Ehrogott, M., Eds.; Springer: New York, NY, USA, 2005; pp. 471–503.
14. Wu, J.Z.; Zhou, Y.P.; Huang, L.; Dong, J.J. Multicriteria correlation preference information (MCCPI) based ordinary capacity identification method. *Mathematics* **2019**, *7*, 300. [CrossRef]
15. Beliakov, G. Construction of aggregation functions from data using linear programming. *Fuzzy Sets Syst.* **2009**, *160*, 65–75. [CrossRef]
16. Wu, J.Z.; Zhang, Q.; Du, Q.; Dong, Z. Compromise principle based methods of identifying capacities in the framework of multicriteria decision analysis. *Fuzzy Sets Syst.* **2014**, *246*, 91–106. [CrossRef]
17. Marichal, J.L.; Roubens, M. Determination of weights of interacting criteria from a reference set. *Eur. J. Oper. Res.* **2000**, *124*, 641–650. [CrossRef]
18. Breiman, L.; Friedman, J.; Stone, C.J.; Olshen, R.A. *Classification and Regression Trees*; CRC Press: Boca Raton, FL, USA, 1984.
19. Loh, W.Y. Classification and regression trees. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2011**, *1*, 14–23.
20. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
21. Resende, P.A.A.; Drummond, A.C. A survey of random forest based methods for intrusion detection systems. *ACM Comput. Surv. (CSUR)* **2018**, *51*, 1–36. [CrossRef]
22. Wu, J.Z.; Beliakov, G. Nonadditivity index and capacity identification method in the context of multicriteria decision making. *Inf. Sci.* **2018**, *467*, 398–406. [CrossRef]
23. Wu, J.Z.; Beliakov, G. Probabilistic bipartition interaction index of multiple decision criteria associated with the nonadditivity of fuzzy measures. *Int. J. Intell. Syst.* **2019**, *34*, 247–270. [CrossRef]

24. Wu, J.Z.; Beliakov, G. Nonmodularity index for capacity identifying with multiple criteria preference information. *Inf. Sci.* **2019**, *492*, 164–180. [CrossRef]
25. Wu, J.Z.; Beliakov, G. Nonadditive robust ordinal regression with nonadditivity index and multiple goal linear programming. *Int. J. Intell. Syst.* **2019**, *34*, 1732–1752. [CrossRef]
26. Beliakov, G.; Wu, J.Z. Learning k-maxitive fuzzy measures from data by mixed integer programming. *Fuzzy Sets Syst.* **2020**. [CrossRef]
27. Marichal, J.L.; Meyer, P.; Roubens, M. Sorting multi-attribute alternatives: The TOMASO method. *Comput. Oper. Res.* **2005**, *32*, 861–877. [CrossRef]
28. Beliakov, G.; Wu, J.Z.; Divakov, D. Towards sophisticated decision models: Nonadditive robust ordinal regression for preference modeling. *Knowl. Based Syst.* **2020**, *190*, 105351. [CrossRef]
29. Huang, L.; Wu, J.Z.; Xi, R.J. Nonadditivity Index Based Quasi-Random Generation of Capacities and Its Application in Comprehensive Decision Aiding. *Mathematics* **2020**, *8*. [CrossRef]
30. Belgiu, M.; Drăguţ, L. Random forest in remote sensing: A review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.* **2016**, *114*, 24–31. [CrossRef]
31. Wu, J.Z.; Xi, R.J.; Zhu, Y. Correlative decision preference information consistency check and comprehensive dominance representation method. *J. Intell. Fuzzy Syst.* **2020**, *38*, 2009–2019. [CrossRef]
32. Wu, J.Z.; Huang, L.; Xi, R.J.; Zhou, Y.P. Multiple Goal Linear Programming-Based Decision Preference Inconsistency Recognition and Adjustment Strategies. *Information* **2019**, *10*, 223. [CrossRef]
33. Grabisch, M. k-order additive discrete fuzzy measures and their representation. *Fuzzy Sets Syst.* **1997**, *92*, 167–189. [CrossRef]
34. Mesiar, R.; Kolesárová, A. k-maxitive aggregation functions. *Fuzzy Sets Syst.* **2018**, *346*, 127–137. [CrossRef]
35. Beliakov, G.; Wu, J.Z. Learning fuzzy measures from data: Simplifications and optimisation strategies. *Inf. Sci.* **2019**, *494*, 100–113. [CrossRef]
36. Wu, J.Z.; Beliakov, G. k-order representative capacity. *J. Intell. Fuzzy Syst.* **2020**, *38*, 3105–3115. [CrossRef]
37. Wang, Z.; Klir, G.J. *Generalized Measure Theory*; Springer Science & Business Media: New York, NY, USA, 2010.
38. Honda, A.; James, S. Parameter learning and applications of the inclusion-exclusion integral for data fusion and analysis. *Inf. Fusion* **2020**, *56*, 28–38. [CrossRef]

© 2020 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/).