

# Mixed Aggregation Functions for Outliers Detection

Hengshan Zhang <sup>a</sup>, Chunru Chen <sup>a,\*</sup>, Tianhua Chen <sup>b</sup>, Zhongmin Wang <sup>a,c</sup>, and Yanping Chen <sup>a,c</sup>

<sup>a</sup>*The School of Computer Science and Technology, Xi'an University of Posts and Telecommunications, Xi'an, China.*

<sup>b</sup>*The Department of Computer Science, School of Computing and Engineering, University of Huddersfield, Huddersfield, United Kingdom.*

<sup>c</sup>*The Shanxi Key Laboratory of Network Data Analysis and Intelligent Processing*

**Abstract.** A scenario that often encounters in the event of aggregating options of different experts for the acquisition of a robust overall consensus is the possible existence of extremely large or small values termed as outliers in this paper, which easily lead to counter-intuitive results in decision aggregation. This paper attempts to devise a novel approach to tackle the consensus outliers especially for non-uniform data, filling the gap in the existing literature. In particular, the concentrate region for a set of non-uniform data is first computed with the proposed searching algorithm such that the domain of aggregation function is partitioned into sub-regions. The aggregation will then operate adaptively with respect to the corresponding sub-regions previously partitioned. Finally, the overall aggregation is operated with a proposed novel consensus measure. To demonstrate the working and efficacy of the proposed approach, several illustrative examples are given in comparison to a number of alternative aggregation functions, with the results achieved being more intuitive and of higher consensus.

Keywords: Aggregation function, concentrate region, t-norm, t-conorm, consensus measure

## 1. Introduction

When dealing with real-world problems, the opinions of different experts are usually aggregated in order to provide more robust solutions. Aggregation functions that aim to combine multiple arguments for the production of a representative consensus are of significant importance in decision making, especially in the presence of the uncertain and noisy settings. It has attracted wide attention for both researchers and practitioners with a number of important approaches and their variations [4,32] proposed in recent literature. Although methods constructed on the basis of fuzzy sets [19] and evidence theory [14] could deal with scenarios under uncertainties, more dedicated approaches are in demand to perform decision aggregation where outliers exist. A common problem that often exists in dealing with decision aggregation is whether the consensus obtained is representative of the majority, reflecting a unbiased universal opinion. It is known that the aggregation of multiple judgements from different sources are often more accurate and reliable

than individuals' dominating judgements. This is also known as the wisdom of crowds [26], with a number of significant applications such as political and economic forecasting, nuclear and chemical safety evaluation and even possible responses to restless volcanoes [26]. A scenario that often encounters when aggregating multiple judgements is the existence of individuals whose judgements may significantly deviate from their peers', which are termed outliers in this paper.

As a numerical example, suppose there are 10 witnesses who give their opinions with respect to the probability of Peter being guilty of a crime as shown in Table 1. Out of 10 different witnesses, 6 tend to believe Peter is more guilty with probability beliefs over 0.5, 1 is totally neutral, while the remaining three opinions favour Peter being innocent. In case of the aggregated result (0.46) calculated over purely arithmetic mean, the final judgement may be counter-intuitive especially considering that the majority vote to the opposite. A conventional approach to improve this situation is to remove the highest and lowest values. For Peter's example again, the probability of being guilty will be 0.481, still leading to a counter-intuitive result, if both the maximal value (0.7) and the minimal value (0.05) are removed.

---

\* Corresponding author. E-mail: chencr9395@163.com.

A very recent review is conducted in [6], which analyses the most popular aggregation operators out of 50 most notable studies. The interrelationship between input argues is typically considered in numerous popular aggregation operators including the discrete Choquet (Ch) integral [15], the Bonferroni Mean (BM) [24], the Muirhead mean [21], the Maclaurin symmetric mean [22] and the Heronian mean (HM) [10]. Additionally, it's also common to observe aggregation operators extended in combination with fuzzy concepts with the aim to tackle uncertainty and imprecision. For instance, the Muirhead mean, which is commonly employed in decision making, is combined with intuitionistic fuzzy numbers in [21], and the Maclaurin symmetric mean has been studied in [22] in combination with linguistic information and intuitionistic fuzzy numbers.

In other literature, the Robust Estimators of Location (REL) [27] is another means to aggregate data with outliers, where the averages are computed using the shortest contiguous sub-sample containing at least half of the data. However, apart from the outliers, non-outliers may also be discarded when aggregation. Another popular set of approaches [30,31] propose to only accept the aggregated result if the consensus measure meets a predefined threshold. However, a possible pitfall of these approaches lies in the mechanism that the original values requires to be modified in order to pass the consensus measure, which may not be desirable for many cases. Alternatively the density-based means (DBM) [2] has been proposed for data with outliers which are weighted by their density. However, in the event of outliers with extremely large or small values, DBM still aggregates counter-intuitive results. Moreover, in the event of no outliers, the aggregated results may also be counter-intuitive, as it may differ from the pure arithmetic mean.

Another interesting and more general type of aggregation operator is the family of Ordered Weighted Averaging (OWA) operators [33], based on which a number of aggregation functions have been proposed [36,35]. OWA is a parameterized operator based on the ordering of extraneous variables that it is applied to. The fundamental aspect of this family of operators is the reordering step in which the extraneous variables are rearranged in descending order, with their values subsequently integrated into a single aggregated one. In particular, the Power Average (PA) function and Power OWA (POWA) function [34] have been proposed, whose weights depend upon the inputs and allow values being aggregated to support and reinforce each other. A number of aggregation functions [32] have been proposed based on the PA function, which allow adjacent values to support and reinforce each other by considering correlations of the underlying data, providing potential benefit to aggregating data with outliers.

The above review suggests that although popular aggregation operators such as the Ch, BM, HM, PA and their extensions take into account the relationship of input arguments, but without explicitly considering scenarios where outliers may exist. While methods such as the REL and DBM may work with outliers,

they could still suffer from undesired pitfalls including discarding non-outliers and generating counter-intuitive results in case of extremely large or small outliers. Inspired by the above observation, this paper proposes a novel approach using a mixed strategy to aggregate consensus where extremely deviated outliers exist. The concentrate region of a set of non-uniform data is first computed such that the domain of aggregation function is partitioned into sub-regions. The aggregation will then operate dynamically with respect to the corresponding sub-regions. In order to demonstrate the working and efficacy of the proposed approach, the illustrated examples are given in comparison to a number of alternative aggregation functions, with results achieved by the proposed method being more intuitive and of higher consensus.

The remainder of this article is organized as follows. In Section 2, the background preliminaries including definition of aggregation function and the methods that have been previously reviewed and later utilised for experimental comparison are introduced. The novel aggregation approach for data with outliers are introduced in Section 3. This is followed by illustrative examples in comparison to several popular state-of-the-arts in Section 4. The comparison with alternative methods is introduced in Section 5. Section 6 provides the conclusion of the paper with future work.

## 2. Preliminaries

In order to demonstrate the proposed approach better, preliminary background is given in this section, which reviews the basic definition of aggregation function as well as a number of related works in dealing with outliers.

### 2.1. Aggregation Function

The aggregation functions that take real arguments from a closed interval  $[a, b] \subset \mathbb{R}$  such as  $[0, 1]$  and produce a real value in  $[0, 1]$  are commonly used in practices. The real values in closed interval  $[a, b]$  and some other types of arguments can be easily transformed to the real values in unit interval. In this article, we conventionally consider the aggregation functions that aggregate the arguments of values taken from the unit interval, which is represented as  $\mathbb{I}$  for simplicity.

**Definition 1 (Aggregation function)** An aggregation function is a function of  $n > 1$  arguments that maps the ( $n$ -dimensional) cube onto an interval  $\mathbb{I} = [0, 1]$ ,  $f : \mathbb{I}^n \rightarrow \mathbb{I}$ , with the properties:

$$1. \underbrace{f(a, a, \dots, a)}_{n\text{-times}} = a \text{ and } \underbrace{f(b, b, \dots, b)}_{n\text{-times}} = b$$

Table 1  
Criminal suspect possibilities given by witnesses

	$W_1$	$W_2$	$W_3$	$W_4$	$W_5$	$W_6$	$W_7$	$W_8$	$W_9$	$W_{10}$	Average
Peter	0.7	0.6	0.7	0.6	0.7	0.6	0.5	0.1	0.05	0.05	0.46

2.  $\mathbf{x} \leq \mathbf{y}$  implies  $f(\mathbf{x}) \leq f(\mathbf{y})$  for all  $\mathbf{x}, \mathbf{y} \in \mathbb{I}$

where  $a, b \in [0, 1]$  are two constants.

The formal definitions of the common aggregations could be found below, which have been proposed for generic decision aggregations, but may not fit for scenarios where extreme outliers exist.

**Definition 2 [4] (Averaging aggregation).** An aggregation function  $f$  is averaging if for every  $\mathbf{X}$  it is bounded by  $\min(\mathbf{X}) \leq f(\mathbf{X}) \leq \max(\mathbf{X})$ , where  $\mathbf{X}$  is a vector and  $\mathbf{X} = \{x_1, x_2, \dots, x_n\}$ .

**Definition 3 [4] (Conjunctive aggregation).** An aggregation function  $f$  has conjunctive behaviour (or is conjunctive) if for every  $\mathbf{X}$  it is bounded by  $f(\mathbf{X}) \leq \min(\mathbf{X}) = \min(x_1, x_2, \dots, x_n)$ .

**Definition 4 [4] (Disjunctive aggregation).** An aggregation function  $f$  has disjunctive behaviour (or is disjunctive) if for every  $\mathbf{X}$  it is bounded by  $f(\mathbf{X}) \geq \max(\mathbf{X}) = \max(x_1, x_2, \dots, x_n)$ .

## 2.2. Robust Estimators of Location

REL delivers values representative of the majority of the data when a specified proportion of data seen as potential outliers is removed [12,27]. In the following, the brief descriptions about the methods including in REL are given. Firstly, the values are ordered, such as  $\mathbf{X} = \{x_{[1]}, x_{[2]}, \dots, x_{[n]}\}$ , where  $x_{[i]}$  denotes the  $i$ th smallest value. Secondly, the averages for the shortest contiguous sub-samples of  $\mathbf{X}$  containing at least half of the values are calculated. The candidate sub-samples are the sets  $\mathbf{X}_k = \{x_{[j]} : j \in \{k, k+1, \dots, k + \lfloor \frac{n}{2} \rfloor\}\}$ ,  $k = 1, 2, \dots, \lfloor \frac{n+1}{2} \rfloor$ . The length of each set is taken as  $\|\mathbf{X}_k\| = |x_{[k + \lfloor \frac{n}{2} \rfloor]} - x_{[k]}|$  and thus the index of the shortest sub-sample is

$$k^* = \arg \min_i \|\mathbf{X}_i\|, i = 1, 2, \dots, \lfloor \frac{n+1}{2} \rfloor$$

The Shorth is the arithmetic mean of  $\mathbf{X}_{k^*}$  and given by

$$\text{Short}(\mathbf{X}) = \frac{1}{h} \sum_{i=1}^h x_i, x_i \in \mathbf{X}_{k^*}, h = \lfloor \frac{n}{2} \rfloor + 1$$

The LMS estimator is the midpoint of  $\mathbf{X}_{k^*}$ . For each sub-sample, the mean is computed  $\bar{x}_k = \frac{1}{h} \sum_{i=1}^h x_i$  and

the corresponding sum of squares is computed  $SQ_k = \frac{1}{h} \sum_{i=1}^h (x_i - \bar{x}_k)^2$ , where  $x_i \in \mathbf{X}_k$ ,  $h = \lfloor \frac{n}{2} \rfloor + 1$ ,  $k = 1, 2, \dots, \lfloor \frac{n+1}{2} \rfloor$ . The LTS solution corresponds to the mean  $\bar{x}_j$  with the smallest associated sum of squares  $SQ_j$ .

$$j = \arg \min_k SQ_k, k = 1, 2, \dots, \lfloor \frac{n+1}{2} \rfloor$$

Let  $S_k = \frac{1}{h} \sum_{i=1}^h |x_i - \bar{x}_k|$ , where  $x_i \in \mathbf{X}_k$ ,  $h = \lfloor \frac{n}{2} \rfloor + 1$ ,  $k = 1, 2, \dots, \lfloor \frac{n+1}{2} \rfloor$ . As,  $\arg \min_k SQ_k = \arg \min_k S_k$  [27], the Least Trimmed Absolute (LTA) deviations estimator is the mean  $\bar{x}_j$  with the smallest associated sum of  $S_j$ .

## 2.3. Density Based Means

DBM is a weighted arithmetic mean and the weights depend on the density of the data [2,5]. The weights of inputs that are closer to the main group of data are higher than the weights of those far away. Let  $d_{ij}$  denote the distance between inputs  $x_i$  and  $x_j$ , the density based mean is defined as follows.

**Definition 5 [5]** The density based mean is the mapping:

$$\text{DBM}(\mathbf{X}) = \sum_{i=1}^n w_i(\mathbf{X})x_i$$

where

$$w_i(\mathbf{X}) = \frac{u_i(x)}{\sum_{j=1}^n u_j(x)} = \frac{K_C(\frac{1}{n} \sum_{j=1}^n d_{ij}^2)}{\sum_{k=1}^n K_C(\frac{1}{n} \sum_{j=1}^n d_{kj}^2)}$$

$d_{ij} = |x_i - x_j|$  and  $K_C$  is the Cauchy kernel given by  $K_C(t) = (1+t)^{-1}$ .

## 2.4. Power Average Aggregation Function

Yager [34] introduced a non linear weighted-average aggregation function, which is called Power Average (PA) aggregation function and can be defined as follows

$$\text{PA}(\mathbf{X}) = \frac{\sum_{i=1}^n (1+T(x_i))x_i}{\sum_{j=1}^n (1+T(x_j))}$$

where  $T(x_i) = \sum_{j=1, j \neq i}^n Sup(x_i, x_j)$  and  $Sup(x_i, x_j)$  denotes the support for  $x_i$  from  $x_j$ , which satisfies the following three properties.

1.  $Sup(x_i, x_j) \in [0, 1]$
2.  $Sup(x_i, x_j) = Sup(x_j, x_i)$
3.  $Sup(x_i, x_j) \geq Sup(x, y)$ , if  $|x_i - x_j| \leq |x - y|$

Obviously, the support (i.e., Sup) measure is essentially a similarity index. The more the similarity, the closer the two values are, and the more they support each other. In this article,  $Sup(x_i, x_j) = 1 - |x_i - x_j|, i \neq j$ .

### 2.5. Discrete Choquet Integral

Choquet Integral (Ch) is a powerful aggregation function especially in merging finite real inputs. Ch focus on capturing the interrelationship among the inputs by adopting strategies to generate the weights of the inputs. The standard Choquet Integral formulas can not accommodate inputs which exist in continuum. So the Discrete Choquet integral based on fuzzy measure  $m$  is proposed [15]. The characteristics of inputs can then be captured by the fuzzy measure, which makes it able to work with outliers. The Discrete Choquet integral is defined as follows

$$Ch(x) = \sum_{i=1}^n x_{(i)} \cdot (m(\{(i), \dots, (n)\}) - m(\{(i+1), \dots, (n)\}))$$

where  $m$  is a monotone set-function satisfying two boundary conditions  $m(\emptyset) = 0$  and  $m(\{1, 2, \dots, n\}) = 1$ .

### 2.6. Bonferroni Mean

Bonferroni Mean (BM) assumes homogeneous relation among the inputs, i.e., each data  $x_i$  is related to the rest of the inputs  $X \setminus \{x_i\}$ , where  $X = \{x_1, x_2, \dots, x_n\}$  denotes the inputs. BM focuses directly on aggregated arguments to capture the interrelationships among them [24]. The degree of interrelationships in between input arguments is measured on the basis of their magnitudes. That is, the interrelationship of input arguments that come with small values tend to be high with respect to those with large values, but low with respect to those also with smaller values. BM is defined as follows <sup>$\frac{1}{p+q}$</sup>

$$BM^{p,q}(a_1, a_2, \dots, a_n) = \left( \frac{1}{n} \sum_{i=1}^n x_i^p \left( \frac{1}{n-1} \sum_{\substack{j=1 \\ i \neq j}}^n x_j^q \right) \right)^{\frac{1}{p+q}}$$

where  $p$  and  $q$  are the parameter and  $p + q > 0$ . Here  $q = p = 1$ .

### 2.7. Heronian Mean

Heronian Mean (HM)[10] is a function that can reflect the interrelationship of the input arguments and is defined as follows:

$$HM(X) = \frac{2}{n(n+1)} \sum_{i=1}^n \sum_{j=i}^n \sqrt{x_i x_j}$$

## 3. Mixed Aggregation Functions Decreasing Effects of Outliers

Despite most of the consensus lies in the so-called Concentrate Region (CR), the information coming from the minority outliers, i.e., the extremely large or small values, should not be simply ignored. In order to tackle this challenge, this paper proposes to utilise a mixed strategy so that appropriate aggregation functions may operate on argument values if they fall into the corresponding sub-regions. In particular, the arguments may be aggregated with averaging function such as arithmetic mean if they fall into the region of CR. Whereas in case of the aggregation for outliers, the key idea is to make the aggregated results as close as those located in the CR, in an attempt to avoid counter-intuitive aggregation behaviour overall.

### 3.1. Algorithm Searching CR for Non-Uniform Data

In order to perform different aggregation strategies depending on the characteristics of the data that may fall into different sub-regions within the universe, it is of significant importance that a method is devised such that it is able to detect the consensus outliers efficiently without human intervention. Instead of assuming any priori distribution of the underlying data which is not always available [16], nor involving high computational calculation by fitting the data into some existing models [18,13], the set of argument values are considered as one-dimensional data, and the sub-region that contains fewer arguments will be iteratively combined into the Concentrate Region if it is closer to the existing Concentrate Region than the data subsets containing large or small values.

Let  $\mathbf{X} = \{x_1, x_2, \dots, x_n\}$  denote the non-uniform data to be aggregated, the average of  $\mathbf{X}$  be  $\bar{x}$ , and the ordered data represented as  $\mathbf{X} = \{x_{[1]}, x_{[2]}, \dots, x_{[n]}\}$ , where  $x_{[i]}$  is the  $i$ th smallest value, with  $x_{[1]} = \min\{\mathbf{X}\}$  and  $x_{[n]} = \max\{\mathbf{X}\}$ . If  $y \notin \mathbf{X}$ , the distance between  $y$  and  $\mathbf{X}$  can be defined as follows:

$$d(y, \mathbf{X}) = \min_{x_i \in \mathbf{X}} d(y, x_i)$$

where  $d(y, x_i) = |y - x_i|$  for computational efficiency.

Three subsets are then initialized as  $S_{min} = \{x_{[1]}\}$ ,  $S_{concentrated} = \{\bar{x}\}$ ,  $S_{max} = \{x_{[n]}\}$ . For the remaining argument values of  $\mathbf{X}$ , each  $x_{[i]}$  is added to the subset, based on the distance between  $x_{[i]}$  and  $S_{min}$ ,  $S_{concentrated}$ ,  $S_{max}$  calculated as *mindistance*  $= \min_k \{d(x_{[i]}, S_k)\}$ , where  $k \in \{min, concentrated, max\}$ . If the distance between  $S_{min}$  and  $S_{concentrated}$  or  $S_{max}$  and  $S_{concentrated}$  are close enough, then they should be combined. The CR is computed according to  $S = S_{concentrated} - \{\bar{x}\}$ ,  $CR = [\min(S), \max(S)]$ . The search algorithm for the acquisition of the CR is demonstrated below in details.

**Step 1** Initialize  $S_{min} = \{x_{[1]}\}$ ,  $S_{concentrated} = \{\bar{x}\}$ ,  $S_{max} = \{x_{[n]}\}$ , and  $i = 2$ ;

**Step 2** Compute the distances between  $x_{[i]}$  and  $S_{min}$ ,  $S_{concentrated}$ ,  $S_{max}$ , respectively, which is presented as  $d(x_{[i]}, S_{min})$ ,  $d(x_{[i]}, S_{concentrated})$  and  $d(x_{[i]}, S_{max})$ ;

**Step 3** Assign  $x_{[i]}$  to the set with minimal distance such that  $x_{[i]} \in S_{k_0}$  given that  $d(x_{[i]}, S_{k_0}) = \min_k \{d(x_{[i]}, S_k)\}$ ,  $k_0, k \in \{min, concentrated, max\}$ ,  $distance[j] = d(x_{[i]}, S_{k_0})$ ,  $j = j + 1$ ;

**Step 4** Update  $i = i + 1$ , if  $i < n$ , go to Step 2, otherwise go to Step 5;

**Step 5** Let  $maxdis = \max_j \{distance[j]\}$ , if  $|\max\{S_{min}\} - \min\{S_{concentrated}\}| \leq maxdis$ ,  $S_{min}$  and  $S_{concentrated}$  are combined with  $S_{concentrated} = S_{min} \cup S_{concentrated}$ . Similarly, if  $|\max\{S_{concentrated}\} - \min\{S_{max}\}| \leq maxdis$ ,  $S_{concentrated}$  and  $S_{max}$  are combined with  $S_{concentrated} = S_{concentrated} \cup S_{max}$ . Finally,  $S = S_{concentrated} - \{\bar{x}\}$  and  $CR = [\min(S), \max(S)]$ .

**Example 1** Let  $\mathbf{X} = \{0.1, 0.2, 0.2, 0.6, 0.6, 0.6, 0.6, 0.7, 0.7, 0.8, 0.97, 0.98\}$  be the arguments to be aggregated.

*Step 1*,  $S_{min} = \{0.1\}$ ,  $S_{concentrated} = \{0.5875\}$ ,  $S_{max} = \{0.98\}$ ,  $i = 2$ .

*Step 2*,  $d(0.2, S_{min}) = 0.1$ ,  $d(0.2, S_{concentrated}) = 0.3875$  and  $d(0.2, S_{max}) = 0.78$ .

*Step 3*,  $d(0.2, S_{min}) = 0.1$  is the minimum, 0.2 is added to  $S_{min}$ ,  $S_{min} = \{0.1, 0.2\}$ ,  $distance[j] = 0.1$ ,  $j = j + 1$ .

*Step 4*,  $i = 3 < 12$ , go to step 2.

Repeat steps 2, 3, 4 until  $i = 4$ ,  $x_i = 0.6$ , go to step 2.

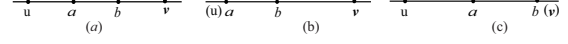


Fig. 1. Figure caption.

*Step 2*,  $d(0.6, S_{min}) = 0.4$ ,  $d(0.6, S_{concentrated}) = 0.0125$  and  $d(0.6, S_{max}) = 0.38$ , where  $S_{min} = \{0.1, 0.2, 0.2\}$  and  $S_{concentrated}, S_{max}$  are not changed.

*Step 3*,  $d(0.6, S_{concentrated}) = 0.0125$  is the minimum, 0.6 is added to  $S_{concentrated}$ ,  $S_{concentrated} = \{0.5875, 0.6\}$ ,  $distance[j] = 0.0125$ ,  $j = j + 1$ .

*Step 4*,  $i = 5 < 12$ , go to step 2.

Repeat steps 2, 3, 4 until  $i = 12$ , go to step 5.

*Step 5*,  $maxdis = \max\{0.1, 0, 0.0125, 0, 0, 0, 0.1, 0, 0.1, 0.01\} = 0.1$ ,  $\max\{S_{min}\} = \max\{0.1, 0.2, 0.2\} = 0.2$ ,  $\min\{S_{max}\} = \min\{0.97, 0.98\} = 0.97$ ,  $\max\{S_{concentrated}\} = \max\{0.5875, 0.6, 0.6, 0.6, 0.6, 0.7, 0.7, 0.8\} = 0.8$ ,  $\min\{S_{concentrated}\} = \min\{0.5875, 0.6, 0.6, 0.6, 0.6, 0.7, 0.7, 0.8\} = 0.5875$ .

As  $|\min\{S_{concentrated}\} - \max\{S_{min}\}| > maxdis$  and  $|\max\{S_{concentrated}\} - \min\{S_{max}\}| > maxdis$ , the Concentrate Region is  $CR = [0.6, 0.8]$ .

### 3.2. Analysing And Classifying Partitioned Sub-Regions of Domains

Based on relative positions between CR and the range of data, there are generally three cases as shown in Fig.1, where  $[a, b]$  is CR, and  $[u, v]$  is the range of data.

In particular, points of CR do not overlap with points of the range of data in Fig.1(a). Whereas, point  $a$  of CR overlaps with the point  $u$  in Fig.1(b), and point  $b$  of CR overlaps with point  $v$  in Fig.1(c). As a result, the partition of the domain may be classified into four cases in the bivariate situation as show in Fig.2, where the axis represents two attributes whose values are combined into a single number.

In particular, case one presented as 1 in Fig.2 is when both values are smaller than CR. Similarly, the second and third cases are when both values are within the range of CR or greater than above CR as shown in 2 and 3 of Fig.2, respectively. In the event of one value is within with the other does not, this is represented as 4-1 in Fig.2. Whereas when one value is above CR and the other is below CR, it is presented as 4-2 in Fig.2. The average of two values cannot be very large or small and may be a value located within CR in the cases 4-1 and 4-2. Therefore, these two cases can be combined into one and it is case 4. In Fig.2, the other

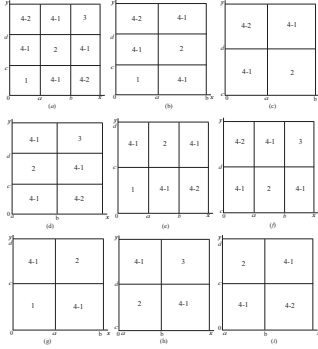


Fig. 2. Classifications for partition of domains

sub-graphs can be considered as the special ones of the sub-graph in Fig.2 (a).

### 3.3. Bivariate Mixed Aggregation Functions Decreasing Effects of Outliers

The behaviours of mixed aggregation functions depend on the inputs, these functions exhibit conjunctive, disjunctive or averaging behaviours on different parts of the domains and there are many different families [4]. Averaging is the most common method to combine information into a single numerical value. The basic rule is that the total score cannot be above or below any of the inputs. The aggregated value is seen as some sort of representative value of all the inputs.

Conjunctive aggregation functions do not allow for compensation: low scores for some criteria cannot be compensated by other scores. The final aggregated result is the smallest value of any of the inputs that bounds the output value. Comparing with conjunctive aggregation functions, disjunctive aggregation functions are the opposite.

The most popular mixed aggregation functions, uninorms [37] and nullnorms [7], are related to triangular norms and conorms [1]. Uninorms make conjunctive operator when dealing with low input values (these below a given value  $e$ ), and have a disjunctive operator for high values (these above  $e$ ) and are averaging for otherwise [4]. Nullnorms are contrary [4]. The characteristics of MADEs (Mixed Aggregation Functions Decreasing Effects of Outliers) were constructed according to Nullnorms.

Without loss of generality, once the partition of domain of the aggregation function done based on the sub-graph shown in Fig. 2 (a), this section discusses the proposed bivariate MADEs. To make the descriptions simpler, let CRs of the aggregated values be the

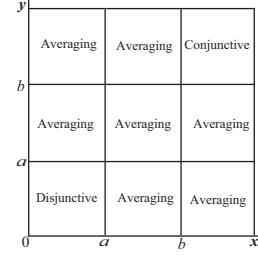


Fig. 3. Behaviors of bivariate MADEs

same in the following definitions, with the behaviours of bivariate MADEs shown in Fig.3.

**Definition 6** Let  $x$  and  $y$  denote the values to be aggregated,  $[a, b] \subseteq [0, 1]$  be concentrate region (CR) for  $x$  and  $y$ . The bivariate MADEs is a mapping,  $DE : [0, 1]^2 \rightarrow [0, 1]$  defined as:

$$DE(x, y) = \begin{cases} \text{Conjunctive} & \min(x, y) > b \\ \text{Disjunctive} & \max(x, y) < a \\ \text{Averaging} & \text{otherwise} \end{cases} \quad (1)$$

where Conjunctive, Disjunctive, Averaging satisfy the following conditions:

$$\begin{aligned} \lim_{x \rightarrow a, y \rightarrow a} \text{Disjunctive}(x, y) &= \lim_{x \rightarrow a, y \rightarrow a} \text{Averaging}(x, y) \\ \lim_{x \rightarrow b, y \rightarrow b} \text{Conjunctive}(x, y) &= \lim_{x \rightarrow b, y \rightarrow b} \text{Averaging}(x, y) \end{aligned}$$

In this paper, the Conjunctive and Disjunctive functions satisfying the above conditions are denoted as  $MC(x, y)$  and  $MD(x, y)$ , respectively. The averaging function usually means arithmetic or geometric mean.

It is then necessary to discuss how to obtain a concrete MADE. Two special classes of conjunctive and disjunctive aggregation functions are the triangular norms and conorms. Hamacher t-norms and t-conorms [4] are the most popular continuous Archimedean t-norm and t-conorm [25] in many practices, with well-known Algebraic and Einstein t-conorms and t-norms as the special cases [28]. For simplicity, Hamacher t-norms and t-conorms were adopted as the special instances in this article, the following Conjunctive and Disjunctive functions are introduced to construct the bivariate MADEs.

**Proposition 1** Based on the Hamacher t-norms and t-conorms, for  $a, b \in (0, 1)$ , the following can be constructed:

$$1) \text{ Let } MD_{\gamma}^H(x, y) = \frac{x + y - \frac{xy}{a} - (1 - \gamma) \frac{xy}{a}}{1 - (1 - \gamma) \frac{xy}{a^2}}, \gamma >$$

$0, x, y \in [0, a], MD_{\gamma}^H$  is monotonously increased about  $\gamma$  and has the following property.

$$\lim_{x \rightarrow a, y \rightarrow a} MD_{\gamma}^H(x, y) = \lim_{x \rightarrow a, y \rightarrow a} \text{Averaging}(x, y)$$

$$2) \text{ Let } MC_\gamma^H(x, y) = \frac{\frac{(x-b)(y-b)}{1-b}}{\gamma + (1-\gamma)\left(\frac{(x-b)}{1-b} + \frac{(y-b)}{1-b} - \frac{(x-b)(y-b)}{(1-b)^2}\right)} +$$

$b, \gamma > 0, x, y \in [b, 1], MC_\gamma^H$  is monotonously decreased about  $\gamma$  and has the following property.

$$\lim_{x \rightarrow b, y \rightarrow b} MC_\gamma^H(x, y) = \lim_{x \rightarrow b, y \rightarrow b} \text{Averaging}(x, y)$$

$$\text{where } \text{Averaging}(x, y) = \frac{x+y}{2}.$$

Based on this proposition, the simple strategy to choose the appropriate functions to construct the MADE can be described as follows. That is, if the effects of extremely small outliers on aggregating result are more significant than the ones with extremely large outliers, for instance, the number of small outliers may be greater than the one with larger outliers or there are no large outliers, then the larger values should be chosen for  $\gamma$ . Otherwise, the smaller values should be chosen for  $\gamma$ . Usually, the averaging aggregation function adopted here is the arithmetic mean. A simplistic example is given below to explain how the bivariate MADEs is constructed.

**Example 2** Let  $[a, b] \subseteq [0, 1]$  be CR of  $x$  and  $y$ . For  $MD_\gamma^H(x, y)$  and  $MC_\gamma^H(x, y)$ , if  $\gamma = 200$ , a useful bivariate MADE (named as bivariate En-MADE) is as follows:

$$En\_DE(x, y) = \begin{cases} \frac{\frac{(x-b)(y-b)}{1-b}}{200 - 199\left(\frac{(x-b)}{1-b} + \frac{(y-b)}{1-b} - \frac{(x-b)(y-b)}{(1-b)^2}\right)} + b & \min(x, y) \geq b \\ \frac{x+y+198\left(\frac{xy}{a^2}\right)}{1+199\left(\frac{xy}{a^2}\right)} & \max(x, y) \leq a \\ \frac{x+y}{2} & \text{otherwise} \end{cases} \quad (2)$$

### 3.4. Multivariate Mixed Aggregation Functions Decreasing Effects of Outliers

Let the aggregated values be a vector  $\mathbf{X} = (x_1, x_2, \dots, x_n)$ , and  $[a, b] \subseteq [0, 1]$  be CR of  $x_i (1 \leq i \leq n)$ . The multivariate MADEs are extended based on previously defined bivariate MADEs, with a set of notations first introduced as below.

1. Let symbolic  $(i)$  represent the index of the  $i$ th largest value and  $[i]$  denote the index of the  $i$ th smallest value for  $\mathbf{X} = (x_1, x_2, \dots, x_n)$ .
2.  $x_i^l$  means  $x_i \leq a$  ( $0 \leq i \leq n_1$ ) and  $n_1$  is the number of values satisfying  $x_i < a$ .
3.  $x_i^m$  represents  $a < x_i < b$  ( $0 \leq i \leq n_2$ ) and  $n_2$  is the number of those values.
4.  $x_i^h$  shows  $x_i \geq b$  ( $0 \leq i \leq n_3$ ) and  $n_3$  denotes the number of values satisfying  $x_i > b$ . In addition,  $n_1 + n_2 + n_3 = n$ .

**Definition 7** Let  $\mathbf{X} = \{x_1, x_2, \dots, x_n\}$  denote the vector of values to be aggregated and  $[a, b] \subseteq [0, 1] (1 \leq i \leq n)$  be CR of  $x_i (1 \leq i \leq n)$ . An Extended Multivariate Disjunctive (EMD) function is a mapping:  $EMD : [0, 1]^n \rightarrow [0, 1]$ . For any  $i (2 < i \leq n)$

$$EMD(X_i) = MD(EMD(x_1, x_2, \dots, x_{i-1}), x_i) \quad (3)$$

where  $X_i = (x_1, x_2, \dots, x_i)$ .

**Definition 8** Let  $\mathbf{X} = \{x_1, x_2, \dots, x_n\}$  denote the vector of values to be aggregated and  $[a, b] \subseteq [0, 1] (1 \leq i \leq n)$  be CR of  $x_i (1 \leq i \leq n)$ . An Extended Multivariate Conjunctive (EMC) function is a mapping:  $EMC : [0, 1]^n \rightarrow [0, 1]$ . For any  $i (2 < i \leq n)$

$$EMC(X_i) = MC(EMC(x_1, x_2, \dots, x_{i-1}), x_i) \quad (4)$$

where  $X_i = (x_1, x_2, \dots, x_i)$ .

**Definition 9** Let  $\mathbf{X} = \{x_1, x_2, \dots, x_n\}$  denote the vector of values to be aggregated and  $[a, b] \subseteq [0, 1] (1 \leq i \leq n)$  be CR of  $x_i (1 \leq i \leq n)$ . The Multivariate MADE is a mapping:  $MADE : [0, 1]^n \rightarrow [0, 1]$ .

$$MADE(X) = M(EMD(x_{(1)}^l, x_{(2)}^l, \dots, x_{(n_1)}^l), \text{Averaging}(x_{(1)}^m, x_{(2)}^m, \dots, x_{(n_2)}^m), EMC(x_{[1]}^h, x_{[2]}^h, \dots, x_{[n_3]}^h)) \quad (5)$$

In the aforementioned definition,  $M$  may be any aggregation functions. However, the main contribution of the proposed MADEs is to decrease the possible side effects of outliers that are located outside of CR. As a result,  $M$  is the weighted arithmetic mean and the weights are defined as the proportions of outliers and the normal argument values are defined as:

$$MADE(X) = \frac{n_1}{n} \times EMD(x_{(1)}^l, x_{(2)}^l, \dots, x_{(n_1)}^l) + \frac{n_2}{n} \times \text{Averaging}(x_{(1)}^m, x_{(2)}^m, \dots, x_{(n_2)}^m) + \frac{n_3}{n} \times EMC(x_{[1]}^h, x_{[2]}^h, \dots, x_{[n_3]}^h) \quad (6)$$

where  $n_1, n_2, n_3$  are numbers of the small outlier values, the values located in CR, and the large outlier values, respectively,  $n_1 + n_2 + n_3 = n$ . If  $\mathbf{W} = \{w_1, w_2, \dots, w_n\} (w_i \in [0, 1])$  is the weight vector corresponding to  $\mathbf{X} = \{x_1, x_2, \dots, x_n\}$ , the weighted aggregation functions for multivariate MADEs are presented as follows

$$MADE(X) = \sum_{(i)=1}^{n_1} w_{(i)}^l \times EMD(x_{(1)}^l, x_{(2)}^l, \dots, x_{(n_1)}^l) + (w_{(1)}^m x_{(1)}^m + w_{(2)}^m x_{(2)}^m + \dots + w_{(n_2)}^m x_{(n_2)}^m) + \sum_{[i]=1}^{n_3} w_{[i]}^h \times EMC(x_{[1]}^h, x_{[2]}^h, \dots, x_{[n_3]}^h) \quad (7)$$

where  $w_{(i)}^l, w_{(i)}^m, w_{[i]}^h$  are the weights corresponding to  $x_{(i)}^l, x_{(i)}^m, x_{[i]}^h$ , respectively. The constructed mixed aggregation function has the following properties.

**Definition 10** Let  $D = [0, 1]^n$  be the definition domain of  $MADE(X)$ , for any  $X = \{x_1, x_2, \dots, x_n\}$ ,  $Y = \{y_1, y_2, \dots, y_n\} \in D$ , the distance is defined as

$$d(X, Y) = \left( \sum_{i=1}^n (x_i - y_i)^2 \right)^{\frac{1}{2}}$$

**Definition 11** Function  $f(x)$  is defined on  $D = [0, 1]^n$ , for any  $X \in D, \forall \varepsilon > 0, \exists \delta > 0$ , if  $d(X, Y) < \delta$ , then  $|f(x) - f(y)| < \varepsilon$ , we define  $f(x)$  is continuous at  $x$ .

**Proposition 2**  $MADE(x)$  is the continuous function about  $x$ .

**Proposition 3** Let  $X, Y \in D = [0, 1]^n$ , if  $X = \{x_{(1)}^t, \dots, x_{(n_1)}^t, x_{(1)}^m, \dots, x_{(n_2)}^m, x_{(1)}^h, \dots, x_{(n_3)}^h\}$ ,  $Y = \{y_{(1)}^t, \dots, y_{(n_1)}^t, y_{(1)}^m, \dots, y_{(n_2)}^m, y_{(1)}^h, \dots, y_{(n_3)}^h\}$ , and  $X \leq Y$  means  $\{x_i < y_i \mid i \in \{1, 2, \dots, n\}\}$ , then  $MADE(X) \leq MADE(Y)$ .

**Example 3** Let  $X = \{x_1, x_2, \dots, x_n\}, [a, b] \subseteq [0, 1](1 \leq i \leq n)$  be CR of  $x_i(1 \leq i \leq n)$ , and  $\gamma = 200$ . We denote  $MD_{200}^H(x_i, x_j) = \frac{x_i + x_j + 198 \frac{x_i + x_j}{a}}{1 + 199 \frac{x_i x_j}{a^2}}$  as  $ES(x_i, x_j)$ , and denote

$$MC_{200}^H(x_i, x_j) = \frac{\frac{(x_i - b)(x_j - b)}{(1 - b)}}{200 - 199 \left( \frac{(x_i - b)}{(1 - b)} + \frac{(x_j - b)}{(1 - b)} - \frac{(x_i - b)(x_j - b)}{(1 - b)^2} \right)} +$$

$b$  as  $ET(x_i, x_j)$ , averaging function is arithmetic mean. Then a special multivariate  $MADE$  can be obtained and it is as follows.

$$MADE_{ES, A, ET}(X) = \frac{n_1}{n} EMD_{ES}(x_{(1)}^l, x_{(2)}^l, \dots, x_{(n_1)}^l) + \frac{n_2}{n} \times \left( \frac{x_{(1)}^m + x_{(2)}^m + \dots + x_{(n_2)}^m}{n_2} \right) + \frac{n_3}{n} EMC_{ET}(x_{[1]}^h, x_{[2]}^h, \dots, x_{[n_2]}^h) \quad (8)$$

One of the biggest advantages of using the proposed searching algorithm lies in the avoidance of manually adjusting associated parameters, which could have made the algorithm sensitive to human errors. For the given one-dimensional data, the distribution of data by searching algorithm is unique, that is, the distribution of data is uniform or the outliers can be found. When aggregating one-dimensional data, the Extended Conjunction function will be used for outliers with larger values. In particular, such function will result in an aggregation, which is less than  $\min \{X_i \mid i = 1, 2, \dots, n\}$  and is more than values lying in the concentrate region, thereby reducing the impact of outliers with high values. Conversely, the application of the extended disjunctive function to outliers with low val-

ues will lead to an aggregated result greater than  $\max \{X_i \mid i = 1, 2, \dots, n\}$  and smaller than any value in the concentrate region.

### 3.5. Consensus Measure for Aggregating Arguments with Outliers

Let  $X \in [0, 1]^n$  be the input vector and  $y$  be the output value. In the case where  $y$  is known, the constructor  $F$  aims to approach  $y$  such that  $F(X) \approx y$ . In the case where  $y$  cannot be obtained in advance, or simply not available, the consensus measure may be utilised to check whether the final result obtained by the aggregation function is consistent with the majority of inputs. For instance, the consensus degree of the preference value may be measured by [29] as  $CD(v_{ij}) = 1 - \frac{1}{m} \sum_{k=1}^m |v_{ij}^{(k)} - v_{ij}^{(c)}|$ , where  $v_{ij}$  is argument to be aggregated and  $m$  denotes the number of experts. Generally speaking, the higher the value, the higher consensus degree it signifies. It is worth noting that the use of consensus measure to gauge the performance of an aggregation function has commonly been adopted in recent literature [17].

In group decision making (GDM), the arithmetic mean of the data located in the concentrate region CR is the value that reflects the closest opinion to the universal one. In order to avoid the overall aggregated result significantly affected by outlier values, it is important for the aggregated results of outliers to approximate the CR mean. A consensus measure is therefore proposed for the final aggregation of argument values with outliers.

Fuzzy implication [23,3] is a function which is monotone non-increasing in the first argument, monotone non-decreasing in the second argument, and satisfies the boundary conditions,  $I(0, 0) = 1, I(1, 1) = 1$  and  $I(1, 0) = 0$ . In this study, the Łukasiewicz implication ( $I_L(x, y) = \min(1, 1 - x + y)$ ) is used to construct the consensus measure.

**Definition 12** Let  $X = \{x_1, x_2, \dots, x_n\}$  denote the argument vector to be aggregated and  $y$  denote the aggregated result and CR of  $x_i$  be  $[a_i, b_i](1 \leq i \leq n)$ . Consensus measure for  $y$  and  $X$  (Presented as  $CRI(X, y)$ ) is defined as:

$$\frac{1}{n} \sum_{i=1}^n \left( \min(1, 1 - (x_i - x_{mid}) + (y - \bar{x}_{CR})) + \min(1, 1 + (x_i - x_{mid}) - (y - \bar{x}_{CR})) - 1 \right) \quad (9)$$



where  $x_{mid}$  is the midpoint of  $\mathbf{X}$ ,  $\bar{x}_{CR} = \frac{1}{n_c} \sum_{j=1}^{n_c} x_j$ ,  $n_c$  is the number of value  $x_j$ , and  $x_j \in [a_j, b_j]$ .

For the above definition,  $|y - \bar{x}_{CR}|$  shows the difference between the aggregated result and the arithmetic mean of the data located in CR. The mid point of  $\mathbf{X}$  essentially discards the outliers with very high and low values,  $|x_i - x_{mid}|$  ( $1 \leq i \leq n$ ) reflects the deviation for  $x_i$  and  $x_{mid}$ , these values are larger for outliers and lower for the normal values. Therefore, the value  $|x_i - x_{mid}|$  ( $1 \leq i \leq n$ ) can approximately show the consensus degree of  $x_i$ . Łukasiewicz implications  $\min(1, 1 - (x_i - x_{mid}) + (y - \bar{x}_{CR}))$  and  $\min(1, 1 + (x_i - x_{mid}) - (y - \bar{x}_{CR}))$  compute the consensus between  $y - \bar{x}_{CR}$  and  $x_i - x_{mid}$  ( $1 \leq i \leq n$ ). In the following, some essential properties for the proposed consensus measure are presented.

1. For  $\mathbf{X} = \{a, a, \dots, a\}$ ,  $a \in [0, 1]$ , it holds that  $CRI(\mathbf{X}, a) = 1$ .
2. For the special case,  $\mathbf{X} = \{0, 0, \dots, 0\}$  and  $y = 1$ , it holds that  $CRI(\mathbf{X}, y) = 0$ .
3. Let  $\pi(i)$  denote all permutations on  $\{1, \dots, n\}$  and  $\mathbf{X} \in [0, 1]^n$ , it holds that  $CRI(x_1, x_2, \dots, x_n) = CRI(x_{\pi(1)}, x_{\pi(2)}, \dots, x_{\pi(n)})$ . Consensus measure CRI is symmetric.

The formula  $\min(1, 1 - (x_i - x_{mid}) + (y - \bar{x}_{CR})) + \min(1, 1 + (x_i - x_{mid}) - (y - \bar{x}_{CR})) - 1$  in definition 12 can measure the consensus degree for each  $x_i$  ( $1 \leq i \leq n$ ) and the aggregating result, this consensus measure is very useful for the data analysis as CRI discards more details in it. More formally, the generalised definition of the consensus measure is as follows.

**Definition 13** Let  $\mathbf{X} = \{x_1, x_2, \dots, x_n\}$  be inputs,  $y$  denote aggregating result, and CR of  $x_i$  be  $[a_i, b_i]$  ( $1 \leq i \leq n$ ). The consensus measure for each Element and the aggregated result (CER) is defined as follows:

$$\min(1, 1 - (x_i - x_{mid}) + (y - \bar{x}_{CR})) + \min(1, 1 + (x_i - x_{mid}) - (y - \bar{x}_{CR})) - 1$$

where  $x_{mid}$  and  $\bar{x}_{CR}$  are same as the ones in definition 12.

According to this definition, CRI is the average of CERs.

#### 4. Illustrative Example

In this subsection, the application is introduced to demonstrate the procedures where MADEs are used to aggregate multiple values and are compared with other methods, such as, Least Squares Trimmed estimator (LTS)[27], Density Based Means (DBM) [2], Power Average aggregation functions (PA) [35], Discret Choquet Integral (Ch) [15], Bonferroni Mean (BM) [24], Heronian Mean(HM)[10] and so on.

The example is on the background of job shop scheduling, where the executive board consisting of a number of experts votes for a certain plan. As it is not a good practice to simply discard or give lowers weights to any one, their votes are usually equally treated.

**Example 4** In an application, the possible plan was produced by using Advanced Planning System (APS) for Flexible Job Shop Scheduling Problem (FJSSP), panel experts will express their degrees of opinions on how much they favour a particular plan, with a example of such probably table given in Table 2. First of all, the concentrate region CR for the set of arguments can be computed resulting in the interval  $[0.6, 0.8]$ . Let  $\gamma = 200$ , EMD and EMC are constructed by using  $MD_\gamma^H$  and  $MC_\gamma^H$  in proposition 1, and the weight of each expert being  $\frac{1}{12}$ . The computing process of MADE is as follows.

$$\begin{aligned} MADE(\mathbf{X}) &= \frac{7}{12} \times (0.8 + 0.6 + 0.7 + 0.6 + 0.7 + 0.6 + \\ &0.6) + \frac{3}{12} \times EMD(0.1, 0.2, 0.2) + \frac{2}{12} \times EMC(0.97, 0.98) \\ &= 0.65714 \times 0.58333 + 0.59971 \times 0.25 + 0.83839 \times \\ &0.16667 = 0.67299 \end{aligned}$$

In comparison to SM, DBM, TM, Mid, PA, LTS, Ch, BM, HM with aggregated results shown in Table 3, it can be observed that MADE computed result (0.67299) is very close to SM (0.65714) than the others computed ones. The CER for each possibility is depicted in Fig.4 in comparison to MADE with TM, DBM, Ch, BM, HM and PA. Although Mid and LTS have higher CRIs than TM as shown in Table 3, they discard too much data including those seen as non-outliers, the comparing results for them with MADE are shown in Fig.5. It can be observed that, for all possibilities, CERs for SM and MADE are very close to each other. Possibilities  $p_i$  ( $1 \leq i \leq 7$ ) are in CR ( $[0.6, 0.8]$ ), CERs for DBM, PA, Ch, HM and TM are less than the ones for MADE. There are same results on possibilities  $p_i$  ( $i = 11, 12$ ), but the results are contrary, for  $p_i$  ( $i = 8, 9, 10$ ). Possibilities  $p_i$  ( $i = 8, 9, 10$ ) have smaller values and are not in CR, the differences between them and SM are larger. MADE computed results are very close to SM and far from  $p_i$  ( $i = 8, 9, 10$ ). CERs between  $p_i$  ( $i = 8, 9, 10$ ) and MADE computed result are lower. DBM, PA, HM and Ch computed results are smaller and far from SM, but MADE computed result is large and very close to

Table 2  
Experts given possibilities on possible plan

	$E_1$	$E_2$	$E_3$	$E_4$	$E_5$	$E_6$	$E_7$	$E_8$	$E_9$	$E_{10}$	$E_{11}$	$E_{12}$
Possibility	0.8	0.6	0.7	0.6	0.7	0.6	0.6	0.1	0.2	0.2	0.97	0.98

Table 3  
Aggregated results and CRI's

	SM	DBM	MADE	TM	Mid	PA	LTS	Ch	BM	HM
results	0.65714	0.59521	0.67299	0.58750	0.60000	0.60327	0.69625	0.58750	0.64880	0.54744
CRI	0.79583	0.76487	0.79319	0.76101	0.76726	0.76890	0.78932	0.76101	0.79166	0.74098

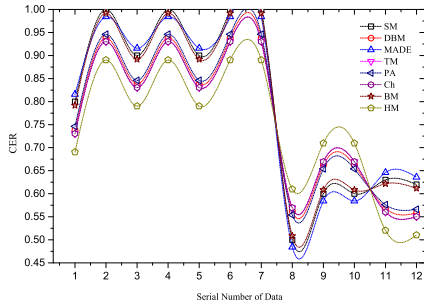


Fig. 4. Comparing results of CERs for MADE with DBM, PA, Ch, BM, HM and TM to possibilities on possible plan

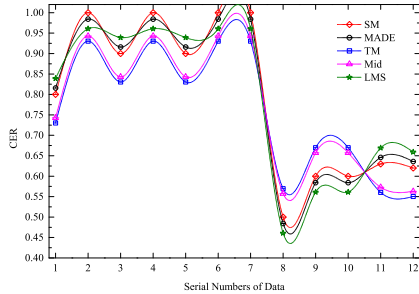


Fig. 5. Comparing results of CERs for MADE with Mid, ITS and TM to possibilities on possible plan

SM. Possibilities  $p_i (i = 11, 12)$  have larger values and are not in CR, CERs for MADE are higher. Together, it can be concluded that CRI of MADE is closer to the one of SM and is higher than the others (Table 3). Although the partial result of MADE is similar to the BM, the CRI of MADE is closer to the one of SM than BM.

In Fig. 5, it can be observed that, for possibilities  $(p_i, 1 \leq i \leq 7)$  which are in CR, CERs for Mid are lower than the ones for MADE. CERs for LTS are lower than the ones for MADE for  $p_2, p_4, p_6, p_7$ , but contrary for  $p_1, p_3, p_5$ . Possibilities  $(p_i, i = 8, 9, 10)$  have smaller values and are not in CR, CERs for

MADE are less than the ones for Mid, but are greater than the ones for LTS. For large values  $(p_i, i = 11, 12)$  that are not in CR, CERs for MADE are greater than the ones for Mid, but less than the CERs for LTS. CERs for 58.33% of possibilities and MADE computed result are higher than ones for these possibilities and LTS obtained result.

In general, CRI for MADE is higher than CRIs for Mid and LTS. The Mid discards most data except the single midpoint, its performance is likely to affect be outliers. The same problems exist for LTS. Based on the discussions given above, it can be concluded that MADE has higher consensus with inputs and it does not discard any input maximally utilising all available information. DBM, PA, Ch, HM and BM do not ignore any argument values, but their consensus degree is lower than that of MADE. SM discards the outliers and is regarded as the standard for comparison. Based on MADE computed result (0.67299), it can be concluded that the APS given plan can be accepted.

The MADE method can also be used in the group decision-making.

**Example 5** Harrisburg and Philadelphia are the famous cities of Pennsylvania in US and they are the alternatives, denoted as  $A = \{A_1, A_2\}$ . The used linguistic term set is  $S = \{s_0 = \text{extremely impossible}, s_1 = \text{less impossible}, s_2 = \text{slight impossible}, s_3 = \text{equally possible}, s_4 = \text{possible}, s_5 = \text{highly possible}, s_6 = \text{extremely possible}\}$ . We invite 10 students to evaluate the possibility of which city is the capital of Pennsylvania, for instance, the opinion of the first student is described as matrix  $E_1$ . It can be translated into a numerical matrix  $T_1$  utilizing the numerical scale model  $p_i = \frac{1}{2}(1 + \log_3^{a_i})$  [38], where  $a_i = (\sqrt{c})^{\Delta^{-1}(s_i)-3}$ ,  $\Delta^{-1}(s_i)=i, c=2$  [39], and  $i$  is the index  $s_i$ .

$$E_1 = \begin{bmatrix} s_3 & s_0 \\ s_6 & s_3 \end{bmatrix} \text{ and } T_1 = \begin{bmatrix} 0.5 & 0.03 \\ 0.97 & 0.5 \end{bmatrix}$$

The possibility of Harrisburg as the capital of Pennsylvania based on the students' opinions can be presented as  $\{0.03, 0.66, 0.82, 0.34, 0.82, 0.82, 0.34, 0.66, 0.97,$

0.66}, and {0.03, 0.97} are the outliers, they cannot be simply discarded. The aggregated result using MADE is presented as matrix,  $R = \begin{bmatrix} 0.5 & 0.61 \\ 0.39 & 0.5 \end{bmatrix}$ . It can be obtained that Harrisburg is the capital of Pennsylvania.

In terms of computation complexity for the proposed method, it is attributed to the computation of its two components. The first step of the proposed method divides the input data into three sub regions based on the distance between data and each sub region, resulting in the computation complexity of  $O(n^2)$ . The second step aggregates values of the three sub regions, resulting in the computation complexity of  $O(n)$ . Overall, the complexity of MADE is therefore  $O(n^2)$ . Although tradition averaging methods such as TM and SM come with  $O(n)$ , they generally cannot work well to generate unbiased result with extreme outliers. Whereas more advanced methods such as DBM, PA, Ch, HM and BM also come with the complexity of  $O(n^2)$ .

## 5. Comparison with Alternative Methods

In summary, integrating the aforementioned analysis, the differences between different methods in this study can be summarized as follows:

1. In Mid and LTS, a fairly large proportion of original argument values may be removed, therefore leading to the conclusion that are not the representative, and their performance of both methods are not stable as shown in Fig. 5 and 1.
2. The outliers which are far from the concentrate area of the data have smaller weights in DBM and PA. Although BM, HM and Ch methods consider the distribution of input data, the overall aggregated result is still not satisfactory as demonstrated by whether Peter is the criminal. In case where outliers are either extremely small or large (instead of a mixture of both), results aggregated by DBM, PA, HM and Ch may still be counter-intuitive, which could be corrected using the proposed MADEs as shown in Section 1.
3. The results aggregated by the proposed method have achieved higher consensus than those of TM, DBM, PA, Ch, HM and BM as demonstrated in Table 3. Moreover, for each aggregated value, MADEs has also achieved higher consensus with the aggregated values that are in the CR, and they have lower consensus with the values that are not in CR as demonstrated in Fig.4 and 5.

## 6. Conclusion

In group decision making, extremely very large or small values that are isolated from others may lead to counter-intuitive aggregated result when data is simply aggregated by averaging. In order to tackle this challenge, this paper proposes a novel approach without discarding any exiting argument value. A search algorithm is first proposed to search the concentrate region for non-uniform data resulting in a number of partitioned domains. Based on the partition of domain, MADEs are constructed and their basic properties are studied. To measure the consensus of aggregated result and the inputs, CRI is proposed, which reflects the consensus degree between aggregated result and the set of all original argument values. Furthermore, CER is proposed, which reflects the consensus degree between the aggregated result and each individual input. Based on computed CERs, more details on consensus between aggregated result and inputs can be obtained. The illustrative examples are given in comparison to a number of alternative aggregation functions, with the results achieved being more intuitive and of higher consensus. Whilst promising results achieved, it is proposed to put this approach into real-world applications such as those from the wisdom of crowds [26]. In addition, the proposed method could be incorporated in conjunction with multi-attribute group decision-making [11], multi-source information fusion [20], fuzzy system construction [9,8].

## 7. Acknowledgement

This work is sponsored by Science and Technology Project in Shaanxi Province of China (Program No.2019ZDLGY07-08), and Xi'an Project for College Talent Providing Services to Enterprise (No. GXYP17.15).

## References

- [1] Claudi Alsina, Berthold Schweizer, and Maurice J Frank, *Associative functions: triangular norms and copulas*. World Scientific, 2006.
- [2] Plamen Angelov and Ronald Yager, Density-based averaging new operator for data fusion. *Information Sciences*, 222:163–174, 2013.
- [3] Michal Baczynski, Gleb Beliakov, Humberto Bustince Sola, and Ana Pradera. *Advances in fuzzy implication functions*. Springer Publishing Company, Incorporated, 2013.
- [4] Gleb Beliakov, Ana Pradera, and Tomasa Calvo. *Aggregation functions: A guide for practitioners*. Springer Publishing Company, Incorporated, 2008.

- [5] Gleb Beliakov, Humberto Bustince Sola, and Tomasa Calvo Snchez. *A Practical Guide to Averaging Functions*. Springer Publishing Company, Incorporated, 2015.
- [6] Fabio Blanco-Mesa, Ernesto León-Castro, and José M. Merigó. A bibliometric analysis of aggregation operators. *Applied Soft Computing*, 81:105488, 2019.
- [7] Tomasa Calvo, Bernard De Baets, and J?nos Fodor. The functional equations of Frank and Alsina for uninorms and nullnorms. *Fuzzy Sets and Systems*, 120(3):385–394, 2001.
- [8] T. Chen, C. Shang, J. Yang, F. Li, and Q. Shen. A new approach for transformation-based fuzzy rule interpolation. *IEEE Transactions on Fuzzy Systems*, pages 1–1, 2019.
- [9] Tianhua Chen, Changjing Shang, Pan Su, and Qiang Shen. Induction of accurate and interpretable fuzzy rules from preliminary crisp representation. *Knowledge-Based Systems*, 146:152–166, 2018.
- [10] Yubao Chen and Peide Liu. Multi-attribute decision-making approach based on intuitionistic trapezoidal fuzzy generalized heronian owa operator. *Journal of Intelligent & Fuzzy Systems*, 27(3):1381–1392, 2014.
- [11] C. Fan, Y. Song, Q. Fu, L. Lei, and X. Wang. New operators for aggregating intuitionistic fuzzy information with their application in decision making. *IEEE Access*, 6:27214–27238, 2018.
- [12] Douglas M. Hawkins and David Olive. Applications and algorithms for least trimmed sum of absolute deviations regression. *Computational Statistics & Data Analysis*, 32(2):119–134, 1999.
- [13] Jinlong Huang, Qingsheng Zhu, Lijun Yang, Dong Dong Cheng, and Quanwang Wu. A novel outlier cluster detection algorithm without top-n parameter. *Knowledge-Based Systems*, 121:32–40, 2017.
- [14] M. A. Islam, D. T. Anderson, A. Pinar, T. C. Havens, G. Scott, and J. M. Keller. Enabling explainable fusion in deep learning with fuzzy integral neural networks. *IEEE Transactions on Fuzzy Systems*, pages 1–1, 2019.
- [15] L. Jin, M. Kalina, R. Mesiar, and S. Borkotokey. Discrete choquet integrals for riemann integrable inputs with some applications. *IEEE Transactions on Fuzzy Systems*, 26(5):3164–3169, Oct 2018.
- [16] Yuh-Jye Lee, Yi-Ren Yeh, and Yu-Chiang Frank Wang. Anomaly detection via online oversampling principal component analysis. *IEEE transactions on knowledge and data engineering*, 25(7):1460–1470, 2013.
- [17] C. Li, Y. Dong, and F. Herrera. A consensus model for large-scale linguistic group decision making with a feedback recommendation based on clustered personalized individual semantics and opposing consensus groups. *IEEE Transactions on Fuzzy Systems*, 27(2):221–233, Feb 2019.
- [18] Sheng Li, Ming Shao, and Yun Fu. Multi-view low-rank analysis with applications to outlier detection. *Acm Transactions on Knowledge Discovery from Data*, 12(3):1–22, 2018.
- [19] J. Liang, R. He, Z. Sun, and T. Tan. Aggregating randomized clustering-promoting invariant projections for domain adaptation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(5):1027–1042, May 2019.
- [20] L. Lin, J. Li, F. Chen, J. Ye, and J. Huai. Road traffic speed prediction: A probabilistic model fusing multi-source data. *IEEE Transactions on Knowledge and Data Engineering*, 30(7):1310–1323, July 2018.
- [21] Peide Liu and Dengfeng Li. Some muirhead mean operators for intuitionistic fuzzy numbers and their applications to group decision making. *Plos One*, 12(1):e0168767, 2017.
- [22] Peide Liu and Xiyu Qin. Maclaurin symmetric mean operators of linguistic intuitionistic fuzzy numbers and their application to multiple-attribute decision-making. *Journal of Experimental & Theoretical Artificial Intelligence*, pages 1–30, 2017.
- [23] M. Mas, M. Monserrat, J. Torrens, and E. Trillas. A survey on fuzzy implication functions. *IEEE Transactions on Fuzzy Systems*, 15(6):1107–1121, 2007.
- [24] A. Mesiarová-Zemánková, S. Kelly, and K. Ahmad. Bonferroni mean with weighted interaction. *IEEE Transactions on Fuzzy Systems*, 26(5):3085–3096, Oct 2018.
- [25] Hung T Nguyen and Elbert A Walker. *A first course in fuzzy logic*. CRC press, 2005.
- [26] Drazen Prelec, H Sebastian Seung, and John McCoy. A solution to the single-question crowd wisdom problem. *Nature*, 541(7638):532–535, 2017.
- [27] P. J. Rousseeuw and A. M. Leroy. *Robust regression and outlier detection*. Wiley-Interscience, 1987.
- [28] Meimei Xia, Zeshui Xu, and Bin Zhu. Some issues on intuitionistic fuzzy aggregation operators based on Archimedean t-conorm and t-norm. *Knowledge-Based Systems*, 31:78–88, 2012.
- [29] Changjiang Xiao, Nengcheng Chen, Chuli Hu, Ke Wang, Zewei Xu, Yaping Cai, Lei Xu, Zeqiang Chen, and Jianya Gong. A spatiotemporal deep learning model for sea surface temperature field prediction using time-series satellite data. *Environmental Modelling & Software*, 120:104502, 2019.
- [30] Zeshui Xu. Attribute weights determination models for consensus maximization in multiple attribute group decision-making. *International Journal of General Systems*, 40(7):755–774, 2011.
- [31] Zeshui Xu. A consensus reaching process under incomplete multiplicative preference relations. *International Journal of General Systems*, 41(4):1–19, 2012.
- [32] Zeshui Xu and Ronald R Yager. Power-geometric operators and their use in group decision making. *IEEE Transactions on Fuzzy Systems*, 18(1):94–105, 2010.
- [33] Ronald R Yager. On ordered weighted averaging aggregation operators in multicriteria decisionmaking , *IEEE Transactions on Systems, Man and Cybernetics*, 18(1):183–190, 1988.
- [34] Ronald R Yager. The power average operator. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, 31(6):724–731, 2001.
- [35] Ronald R Yager. Generalized OWA aggregation operators. *Fuzzy Optimization and Decision Making*, 3(1):93–107, 2004.
- [36] Ronald R Yager and Dimitar P Filev. Induced ordered weighted averaging operators. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 29(2):141–150, 1999.
- [37] Ronald R Yager and Alexander Rybalov. Uninorm aggregation operators. *Fuzzy sets and systems*, 80(1):111–120, 1996.
- [38] Zhang H, Zheng Q, Wang Z, et al. Crowd intelligence for decision making based on positive and negative comparing with linguistic scale. 2018 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), IEEE, 2018: 1-8.
- [39] Dong Y, Hong W C, Xu Y, et al. Selecting the individual numerical scale and prioritization method in the analytic hierarchy process: A 2-tuple fuzzy linguistic approach. *IEEE Transactions on Fuzzy Systems*, 19(1): 13-25, 2010.