Professor Jian-Zhang WU, PhD
E-mail: sjzwjz@gmail.com；
School of Business, Ningbo University, China
Yan-Qing LI, MS Candidate
E-mail: hhuliyanqing@163.com;
School of Business, Ningbo University, Ningbo, China
Feng-Feng CHEN, MS Candidate
E-mail: chenfengfengg@163.com;
School of Business, Ningbo University, Ningbo, China
Professor Gleb BELIAKOV, PhD
E-mail: gleb@deakin.edu.au;
School of Information Technology, Deakin University
Melbourne, Australia

SELF-ORGANIZED AND UNSUPERVISED INFORMATION INTEGRATION METHOD BY COMBINING CAPACITY AND PCA

Abstract. The Choquet capacity and integral is a widely accepted model to aggregate the multiple dimensional information mainly for its preeminent ability to flexibly represent and efficiently deal with the interaction phenomenon among the multiple correlative criteria/items. The PCA (Principal Component Analysis) is a popular approach to obtain the successive importance of multiple criteria as well as the correlations among them from the inter-correlated multiple dimensional data. In this paper, we combine some unsupervised capacity identification methods and PCA to constitute a self-organized information aggregation scheme, in which the PCA is used to generate the partial or comprehensive importance and correlations of the reduced (if necessary) multiple criteria, then adopting these preference information as the input, the capacity identification methods transform them into the competent capacities according to some rules and principles, and the final comprehensive integrated information is carried out by the Choquet integral accordingly. The main characteristic of this scheme is it only takes the raw unscaled data on the multiple criteria as the input and many meaningful aggregation analysis and decision aid results can be automatically output without the analyst or decision maker's supervision. The feasibility and practicability of the proposed models are demonstrated by using the QS World University Rankings data.
Keywords: Capacity, PCA, MCCPI, Information Aggregation, self-organized.

Jian-Zhang Wu, Yan-Qing Li, Feng-Feng Chen, Gleb Beliakov

_____

1. Introduction

The Choquet capacity [4], also called fuzzy measure [1, 15], or nonadditive measure [18], extends the probability measure by replacing the additivity with respect to disjoint subsets by the monotonicity with respect to set inclusion, whose explicit expression is the nonadditivity with respect to disjoint subsets. With the monotonicity or nonadditivity property, the capacity becomes an excellent tool to flexibly describe a variety of interaction situations of multiple dimensions, e.g., the multiple decision criteria or evaluation items [19, 20]. The Choquet integral [4] w.r.t. capacity has many good aggregation characteristics and is the most common type of fuzzy or nonlinear integral to aggregate the correlated partial evaluation or performance information on the reciprocally interacting criteria/items [9].

One of the major tasks of the Choquet capacity and integral information integration scheme is to identify the competent capacity for representing the given implicit or explicit preference information. The main capacity identification methods can be classified into two categories.

The first category of capacity identification methods aim to fit the learning set, which consist of a number of history aggregation instance's partial evaluations as well as their overall evaluation or ranking orders, usually regarded as implicit preference information [23]. Their general idea is to find out the most suitable capacity that leads to the least divergence between the obtained overall evaluations and the desired ones given in the learn set by using some linear or nonlinear optimization models [7]. For example, the least-squares principle [7], the least absolute deviation criterion [3] and the maximum log-likelihood principle [5] try to fit the desired overall evaluations of given instances; meanwhile the maximum split approaches [14] try to fit the desired ranking order of given instances. This category can be regarded as the typical supervised machine learning method [7].

In addition, some other supervising preference information about the importance or interaction of the criteria/items can be further added into the capacity identification models, like one criterion is more important than another one and interaction of some criteria combination is strongly positive/negative [1]. This type of information is usually called the explicit preference information [22], which can be represented by the comparison and interval range forms of certain equivalent representation, i.e., a linear one-to-one mapping in general, such as the Möbius representation, Shapley simultaneous interaction index[6], nonadditivity index [18, 20] and nonmodularity index [19].

The second category of capacity identification method only use the explicit preference information to constitute the compatible constraints [21, 24] as well as the objective function of the optimization model and without resorting to the learning set. These methods mainly consist of the maximum entropy principle method [12], the compromise principle method [23] and the MCCPI (multiple criteria correlation preference information) based minimum deviation method [22]. The maximum entropy principle, or equally the principle of minimum distance

_____

with symmetric additive capacity [1], pursues the capacity with the maximum entropy as the most desired one, which basically tries at most to ensure each criterion/item plays the same effect on the overall evaluations. By contrast, the compromise principle method manages to ensure the chance parity of all the instances or candidates to obtain their maximum evaluations [23]. The MCCPI method aims to find out the optimal capacities that have the minimum deviation with the initial pairwise preference information.

Only with the input preference information of the multiple items and partial evaluations of the instances, the second category of identification methods can automatically generate the most satisfied capacity and then by using the Choquet integral the final aggregation information or overall evaluation can be obtained in sequence. More specially, the maximum entropy principle method only needs the comparison and interval range forms of the preference information; beside this type of information, the compromise method also needs the partial evaluations of all instances; the MCCPI method needs a special pairwise comparison of all criteria/items by the aid of a specific two dimensional diagram or scale table. Traditionally, these input information is provided by the decision maker or data analyst which is just kind of his/her subjective judgment on the importance and interaction of these evaluation criteria/items.

However, in some situation especially for facing the rather new and very complex problem, the decision maker or analyst can not provide high quality or reliable input preference information since the lack of background knowledge or the complexity and uncertainty inherent in the information aggregation problem. Hence, the self-organized and unsupervised method to automatically integrate the correlated information becomes very useful, at least it can provide certain aid analyses on the matter or give some decision supports about the issue.

Fortunately, the PCA (principal component analysis) method is a wide used tool to represent the inter correlated data into new orthogonal variables with successively maximize variance [17], which is generally accepted as a good tool to reduce dimensionality [10] and some useful information of the importance and correlation of the initial items can be derived accordingly. And by adopting some suitable transformation rules, these importance and correlation information generated by PCA can be smoothly transformed into the required input preference information for the second category capacity identification methods. In this paper, we combine the PCA with the maximum entropy method, the compromise principle method, the MCCPI method and a further proposed new type of capacity identification method, to totally construct four types of self-organized and unsupervised multiple dimensions information aggregation approaches.

This paper is organized as follows. After the introduction, we present some knowledge about the capacity, Shapley simultaneous interaction index and the Choquet integral in Section 2. In Sections 3 and 4, we briefly introduce the relative capacity identification methods and PCA, respectively. Section 5 is devoted to

Jian-Zhang Wu, Yan-Qing Li, Feng-Feng Chen, Gleb Beliakov

_____

combination schemes and rules of the PCA and capacity identification methods, a new identification model is proposed as well. We demonstrate the steps and applicability of the proposed unsupervised information aggregation method by using the QS World University Rankings data 2019 edition and some empirical comparison analyses are also given. Finally, we conclude the paper in Section 6.

2. Preliminaries

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

Definition 1 *[1, 4] A capacity on N is a set function $\mu: \mathcal{P}(N) \to [0,1]$ with (i) $\mu(\emptyset) = 0$, $\mu(N) = 1$; (ii) $\forall A, B \subseteq N$, $A \subseteq B$ implies $\mu(A) \leq \mu(B)$.*

The quantity of $\mu(A)$ represents the importance of this criteria coalition to information aggregation problem. The monotonicity w.r.t. inclusion subsets also enables the capacity transcend the probability measure whose main characteristic is the additivity w.r.t. disjoint subsets to flexible represent the interaction or dependency among multiple correlated criteria or items [1]. The interaction kind and density is usually explicitly represented by certain interaction index, e.g., the Shapley simultaneous interaction index given below.

Definition 2 *[8] Let μ be a capacity on N, the Shapley importance and interaction index of μ is defined as, for any $A \subseteq N$,*

$$I_\mu(A) = \sum_{B \subseteq N \setminus A} \frac{1}{(|N| - |A| + 1)} \binom{|N| - |A|}{|B|}^{-1} \left(\sum_{C \subseteq A} (-1)^{|A \setminus C|} \mu(C \cup B) \right).$$

Generally speaking, $I_\mu(i)$ is regarded as the overall importance of criterion $i \in N$, and $I_\mu(A)$, $|A| \leq 2$, the comprehensive simultaneous interaction of subset A. For the singletons and 2-order subsets, we have [6, 8]

$$\sum_{i \in N} I_\mu(\{i\}) = 1, \ I_\mu(\{i\}) \geq 0, \text{ and } I_\mu(\{i, j\}) \in [-1, 1], \ i, j \in N.$$

Remark 1. *Actually, there are many other interaction indices, like nonadditivity index [18], nonmodularity index [19] and bipartition interaction indices [20]. Different types of indices have some different properties and practical meanings for the aggregation process. For simplication, we only introduce the Shapley interaction index.*

The Choquet integral is a widely accepted form of fuzzy integral to flexibly aggregate the partial evaluation information of instances or candidates on multiple interdependent criteria/items.

Definition 3 *[16] Let x be a real-valued function on N, $x := (x_1, \ldots, x_n) \in (-\infty, \infty)^n$, the Choquet integral of x with respect to a capacity μ on N is defined as*

_____

$$C_\mu(x) = \sum_{i=1}^{n} [x_{(i)} - x_{(i-1)}]\mu(N_{(i)}) = \sum_{i=1}^{n} [\mu(N_{(i)}) - \mu(N_{(i+1)})]x_{(i)}$$

where the parentheses used for indices represent a permutation on N such that
$x_{(1)} \leq \cdots \leq x_{(n)}$, $x_{(0)} = 0$, $N_{(i)} = \{(i), \ldots, (n)\}$, and $N_{(n+1)} = \emptyset$.

3. The explicit preference information based capacity identification methods

The most common type of the explicit preference information is the comparison and range interval forms of the importance and interaction of criteria/items [1, 7], like the overall importance of criterion i is at least as great as that of criterion j: $I_\mu(\{i\}) - I_\mu(\{j\}) \leq 0$; the sign of comprehensive interaction of two criteria $\{i, j\}$ is positive or negative: $I_\mu(\{i, j\}) \in [\delta, 1]$ or $I_\mu(\{i, j\}) \in [-1, -\delta]$, $\delta > 0$; the comprehensive interaction of criteria subset A is larger than that of B with threshold $\delta > 0$: $I_\mu(A) - I_\mu(B) \geq \delta$. Usually the preference information are linear constraints, if they are compatible reciprocally, a feasible range can be constituted. Otherwise, some inconsistency recognition and adjustment methods can be adopted to change them into the consistent case, see [21].

3.1. The maximum entropy principle method

The entropy of a capacity μ on N is defined by [13]

$$E(\mu) = \sum_{i=1}^{n} \sum_{A \subseteq N \setminus \{i\}} \frac{(|N| - |A| - 1)! |A|!}{|N|!} h(\mu(A \cup \{i\}) - \mu(A)), \tag{1}$$

where $h(x) = -x \ln x$ if $x > 0$ and 0 if $x = 0$. Since the symmetric additive capacity has the largest entropy value $\ln(n)$ [2, 23], we can alternatively calculate the minimum distance with the symmetric additive capacity to realize the maximum entropy principle, e.g., minimum variance principle [12]. If adopt the absolute distance, we can get the following multiple goal linear programming (MGLP) :

$$\min z = \sum_{A \subset N} d_A^+ + d_A^-$$
$$s.t. \begin{cases} \text{The boundary and monotonicity condition of capacity,} \\ \text{The explicit preference information,} \\ \mu(A) - d_A^+ + d_A^- = |A|/|N|, d_A^+, d_A^- 0, \forall A \subset N, \end{cases} \tag{2}$$

where d_A^+, d_A^- are the positive and negative deviation variables holding the relation $d_A^+ \times d_A^- = 0$. Since the objective is minimum, the nonlinear constraints $d_A^+ \times d_A^- = 0$ can be removed totally, then the above model remains a linear programming.

3.2. The compromise principle method

As a kind of contrast, the compromise principle method [23] tends to give each instance or candidate the equal chance to reach the maximum overall evaluation. Suppose the instance set is $X = \{x^k, k = 1, \ldots, |X|\}$, stance $x^k := (x_1^k, \ldots, x_n^k) \in (-\infty, \infty)^n$ and the maximum reachable overall evaluation of x^k is denoted as $\max x^k$, then the compromise principle can be realized by the following MGLP model:

65

Jian-Zhang Wu, Yan-Qing Li, Feng-Feng Chen, Gleb Beliakov

_____

$$\min z = \sum_{k=1}^{|X|} d_k^+ + d_k^-$$

$$s.t. \begin{cases} \text{The boundary and monotonicity condition of capacity,} \\ \text{The explicit preference information,} \\ C(x^k) - d_k^+ + d_k^- = \max x^k, d_k^+, d_k^- \geq 0, k = 1, \ldots, |X|, \end{cases} \tag{3}$$

where d_k^+, d_k^- are the positive and negative deviation variables. In fact, we can just set $\max x^k$ as $\max(x_1^k, \ldots, x_n^k)$ for convenience.

3.3 The MCCPI based minimum deviation method

For each pair of items $i, j \in N$, if use I_i^{ij} to represent the relative importance of item i with the comparison to item j, I_{ij}^P the estimated value of the partial interaction between i and j without considering other items, we can have $\binom{n}{2}$ triple-element vectors $(I_i^{ij}, I_j^{ij}, I_{ij}^P), i, j \in N, i < j$, called the MCCPI for simplicity, where

$$0 \leq I_i^{ij}, I_j^{ij} \leq 1, \ I_i^{ij} + I_j^{ij} = 1, \ |I_{ij}^P| \leq 2\min(I_i^{ij}, I_j^{ij}) \tag{4}$$

The following equations bridge the MCCPI to the Shapley interaction index of the most desired capacity [22, 24]:

$$I_i^{ij} = \frac{I_\mu(\{i\})}{I_\mu(\{i\}) + I_\mu(\{j\})}, I_{ij}^P = \frac{I_\mu(\{i,j\})}{I_\mu(\{i\}) + I_\mu(\{j\})}. \tag{5}$$

By introducing the absolute distance, we can realize the minimum deviation principle between the left and right hand sides of the above two equations by the following MGLP model:

$$\min z = \sum_{i,j \in N, i<j} d_{ij}^+ + d_{ij}^- + e_{ij}^+ + e_{ij}^-$$

$$s.t. \begin{cases} \text{The boundary and monotonicity condition of capacity,} \\ I_i^{ij}(I_\mu(\{i\}) + I_\mu(\{j\})) - I_\mu(i) - d_{ij}^+ + d_{ij}^- = 0, i, j \in N, i < j, \\ I_{ij}^P(I_\mu(\{i\}) + I_\mu(\{j\})) - I_\mu(\{i,j\}) - e_{ij}^+ + e_{ij}^- = 0, i, j \in N, i < j, \end{cases} \tag{6}$$

where $d_{ij}^+, d_{ij}^-, e_{ij}^+, e_{ij}^- \geq 0$ are the positive and negative deviation variables.

4. The PCA and preference information generation

The PCA allows us to elicit some useful information from multiple inter-correlated quantitative items or dimensions. It represents the original data into new inter-orthogonal variables, which are basically linear combinations of original items. According to the contribution rates to total variance, some of the new variables are selected as the principle components to realize the aim of dimension reduction [10].

Suppose there is a data set $\mathbf{S} = [s_{ki}]_{m \times n}$ with m instances and n inter-correlated items or dimensions. First, it is better to normalize each column of data $\mathbf{S}$ into the same scale, generally with standard deviation one and mean zero. Denote the normalized data as $\overline{S} = [\overline{s}_{ki}]_{m \times n}$, $\overline{s}_{ki}$ is given by

_____

$$\overline{s}_{ki} = \frac{s_{ki} - mean(s_i)}{sd(s_i)}, \tag{7}$$

where s_i is the i column of the data $\mathbf{S}$, $mean(s_i)$ is the mean of s_i and $sd(s_i)$ is the standard deviation of s_i. Then, based on the covariance matrix of $\overline{S}$, PCA can generate the n orthogonal new variables:

$$V_i = v_{i1}\overline{s}_1 + \cdots + v_{ii}\overline{s}_i + \cdots + v_{in}\overline{s}_n, i = 1, \ldots, n, \tag{8}$$

as well as their corresponding eigenvalues, denoted as λ_i, $i = 1, \ldots, n$, $\lambda_1 > \lambda_2 > \cdots > \lambda_n$. λ_i shows the contribution of V_i to the total variance and is used to determine whether V_i can be a reserved principle components or not. Finally, suppose only $n_0 n$ new variables are chosen as the principle components, we can obtain some information about the importance and correlation among the original items or dimensions.

Based on the result of PCA, the importance of items i, denoted as RI_i, can be **partially** estimated by:

$$PI_i = \sum_{l=1}^{n_0} \lambda_l |v_{li}| \tag{9}$$

and the **comprehensive/overall** importance of item i can be estimated by:

$$CI_i = \frac{\sum_{l=1}^{n_0} \lambda_l |v_{li}|}{\sum_{i=1}^{n} \sum_{l=1}^{n_0} \lambda_l |v_{li}|}. \tag{10}$$

For the interaction between two original items i, j, it can be **partially** estimated from the correlation coefficients of the scaled data $\overline{S}$ as:

$$PC_{ij} = \frac{cov(\overline{s}_i, \overline{s}_j)}{sd(\overline{s}_i) sd(\overline{s}_j)}, \tag{11}$$

and their **comprehensive** correlation coefficient can be given as:

$$CC_{ij} = \frac{cov((v_{i1}, \ldots, v_{in_0}), (v_{j1}, \ldots, v_{jn_0}))}{sd(v_{i1}, \ldots, v_{in_0}) sd(v_{j1}, \ldots, v_{jn_0})}, \tag{12}$$

where $cov()$ is the covariance function and $sd()$ is the standard deviation function, hence the correlation coefficients are basically the Pearson correlation coefficients.

The partial/comprehensive importance and correlation coefficients, PI_i, CI_i, PC_{ij} and CC_{ij}, are vital references to generate the input preference information of the capacity identification methods.

Remark 2. *The partial importance PI_i is not a normalized value, i,e., not always less than or equal to 1. If we want to reduce the initial items, we can use PI_i or CI_i as an indicator to remain the given number items with the largest estimated importance. For the correlation coefficient, PC_{ij} is obtained only by the involved two columns of data and taken as the partial correlation, meanwhile CC_{ij} is obtained by the coefficients of the orthogonal variables constituted by all the items, which are also used by most of the software packages for PCA [11], and hence can be taken as the comprehensive correlation coefficient.*

Jian-Zhang Wu, Yan-Qing Li, Feng-Feng Chen, Gleb Beliakov
_____

4.1. Transformation rules to comparison and interval preference

The maximum entropy principle method as well as compromise principle method needs the comparison and interval forms of preference information as their input information.

For the importance of items, we can first sort all the comprehensive importance CI_i, $i = 1, \ldots, n$, in an descending order:

$$CI_{[1]} > CI_{[2]} > \cdots > CI_{[n]}, \tag{13}$$

where $[\cdot]$ means a permutation on N, then transform it into a series of preference constraints of Shapley importance indices:

$$
\begin{aligned}
I_\mu(\{[1]\}) - I_\mu(\{[2]\}) &\geq \delta_1, \\
I_\mu(\{[2]\}) - I_\mu(\{[3]\}) &\geq \delta_1, \\
&\ldots \\
I_\mu(\{[n-1]\}) - I_\mu(\{[n]\}) &\geq \delta_1,
\end{aligned}
\tag{14}
$$

where $\delta_1 > 0$ is the threshold. For the interactions of items, we can simply classify them into positive and negative ones as:

$$
\begin{aligned}
I_\mu(\{i,j\}) \in [\delta_2, 1] &\quad \text{if} \quad CC_{ij} < 0, \\
I_\mu(\{i,j\}) \in [-1, -\delta_2] &\quad \text{if} \quad CC_{ij} > 0,
\end{aligned}
\tag{15}
$$

where $\delta_2 > 0$ is the threshold.

Remark 3. *Empirically, we can set the threshold in Eq. (14) as $\delta_1 = min(0.05, 1/(4n))$, which aims to give each index a reasonable feasible range. For Eq. (15), we can just set $\delta_2 = 0.05$ to differentiate the index with zero. It also needs to point out that the Eq. (15) shows that the correlation coefficient is kind of opposite concept of interaction. Actually, it is reasonable that if two items are redundant/complementary (i.e., their interaction is negative/positive), then their correlation coefficient is likely to be positive/negative (i.e., the data of two items may highly be with similar/contrary trend).*

Remark 4. *It is infrequent to encounter that two comprehensive importance are equal. In that case, we can transform $RI_{[i]} = RI_{[j]}$ into two constraints: $RI_{[i]} - RI_{[j]} \leq \delta_1$ and $RI_{[i]} - RI_{[j]} \geq -\delta_1$. Furthermore, if the comprehensive correlation coefficient is equal to zero, we can transform $CC_{ij} = 0$ into two constraints: $CC_{ij} \leq \delta_2$ and $CC_{ij} \geq -\delta_2$.*

The Eqs. (14) and (15) can be directly adopted in the maximum entropy and compromise principle methods as the hard constraints. One can see that, in Eq. (15), we only use the kind (positivity and negativity) of the correlation coefficients to get the relatively rough intervals of the interaction indices, mainly because many refined intervals mean many hard constraints and will dramatically increase the infeasibility chance of the model and then bring tremendous efforts on inconsistency recognition and adjustment. To avoid this situation, in the following two subsections, we give some MGLP based models and algorithms to more detailedly use the values of the partial and comprehensive correlation coefficients.

_____

4.2. Transformation rules to MCCPI information

The MCCPI preference information, $(I_i^{ij}, I_j^{ij}, I_{ij}^P), i, j \in N, i < j$, is obtained from the pairwise comparison and can be regarded as a sort of partial information. Hence, we can get their preference information from the partial importance and partial correlation coefficients as:

$$
\begin{aligned}
I_i^{ij} &= \frac{PI_i}{PI_i + PI_j}, \\
I_j^{ij} &= 1 - I_i^{ij} = \frac{PI_j}{PI_i + PI_j}, \\
I_{ij}^P &= -PC_{ij}.
\end{aligned}
\tag{16}
$$

Then construct and solve the MCCPI model in Eq. (6), we can get a competent capacity.

In order to observe more aggregation results to support a comprehensive and credible analysis or decision, we can adopt the idea of random generation MCCPI given in [24] and introduce two positive variables δ_3 and δ_4 to define the following allowable ranges:

$$
\begin{aligned}
I_i^{ij} &\in \left[\frac{PI_i}{PI_i + PI_j} - \delta_3, \frac{PI_i}{PI_i + PI_j} + \delta_3 \right], \\
I_{ij}^P &\in [-PC_{ij} - \delta_4, -PC_{ij} + \delta_4,],
\end{aligned}
\tag{17}
$$

and $I_j^{ij} = 1 - I_i^{ij}$ accordingly, where two thresholds can be set as $\delta_3 = 0.05$ and $\delta_4 = 0.1$ for example. The following algorithm 1 uses the random MCCPI in the allowable ranges to generate more desired capacities and the corresponding comprehensive aggregation results.

Algorithm 1: Random MCCPI based information aggregation.

Initialize: the optimal capacities matrix $\mathbf{U}_{\mathcal{K} \times 2^{|N|}}$, where $\mathcal{K}$ is the amount of desired capacities, the overall evaluations matrix $\mathbf{E}_{\mathcal{K} \times |X|}$.

for k *in* $1 : \mathcal{K}$ **do**

    for *each pair of items* $\{i, j\} \subseteq N$ **do**

      | get a random point in their MCCPI's allowable range (see (17)).

    end

    Solve model (6) with all the selected MCCPI and obtain thee optimal capacity, μ_k, and store it as kth row of $\mathbf{U}$,

    Calculate the Choquet integrals of all instances w.r.t. μ_k, and store it as kth row of $\mathbf{E}$.

end

Output: the matrices $\mathbf{U}$ and $\mathbf{E}$.

Basically, in this subsection, we just transform the partial preference information into the goal constraints as well as the multiple goal objective function and use the MGLP based MCCPI method to obtain some competent capacities. In the next subsection, the comprehensive preference information is adopted to identify the desired capacities.

_____

4.3. Comprehensive importance and interaction oriented capacity identification method

Actually, the comprehensive importance and correlation coefficients in Eqs. (10) and (12) can be taken as the references of Shapely importance and interaction indices of the items:

$$
\begin{aligned}
I_\mu(\{i\}) &= CI_i, i \in N, \\
I_\mu(\{i,j\}) &= -CC_{ij}, i,j \in N, i < j.
\end{aligned}
\tag{18}
$$

Then we can use the following MGLP to get a desired capacity:

$$
\min z = \sum_{A \subset N, |A|=1,2} d_A^+ + d_A^-
$$

$$
s.t. \quad \begin{cases} \text{The boundary and monotonicity condition of capacity,} \\ I_\mu(\{i\}) - CI_i - d_i^+ + d_i^- = 0, i \in N, \\ I_\mu(\{i,j\}) + CC_{ij} - d_{ij}^+ + d_{ij}^- = 0, i,j \in N, i < j, \end{cases}
\tag{19}
$$

where d^+, d^- are the positive and negative deviation variables.

Similarly, we can also further introduce two positive variables δ_5 and δ_6 to get the following allowable ranges:

$$
\begin{aligned}
I_\mu(\{i\}) &\in [CI_i - \delta_5, CI_i + \delta_5], i \in N, \\
I_\mu(\{i,j\}) &\in [-CC_{ij} - \delta_6, -CC_{ij} + \delta_6], i,j \in N, i < j.
\end{aligned}
\tag{20}
$$

The algorithm 2 just uses the random comprehensive preference information in the above allowable ranges to generate the competent capacities.

Algorithm 2: Random comprehensive preference based information aggregation.

Initialize: the optimal capacities matrix $\mathbf{U}_{\mathcal{K} \times 2^{|N|}}$, where $\mathcal{K}$ is the amount of desired capacities, the overall evaluations matrix $\mathbf{E}_{\mathcal{K} \times |X|}$.

for k in $1 : \mathcal{K}$ **do**
  for $each$ $item$ i of $decision$ $criteria$ $\{i,j\} \subseteq N$ **do**
    get a random point in its Shapley importance index's allowable range (see (20)).
  end
  for $each$ $pair$ of $items$ $\{i,j\} \subseteq N$ **do**
    get a random point in their Shapley interaction index's allowable range (see (20)).
  end
  Construct and solve the MGLP accoridng to Eq. (19), denote its optimal capacity as μ_k, and store it as kth row of $\mathbf{U}$,
  Get the Choquet integrals of all instances w.r.t. μ_k, and store it as kth row of $\mathbf{E}$.
end
Output: the matrices $\mathbf{U}$ and $\mathbf{E}$

---

5. An illustrative application in ranking top 100 universities

The QS World University Rankings comprehensively compare and rank the universities all over the world according to the following six criteria: (1) Academic Reputation, (2) Employer Reputation, (3) Faculty Student, (4)International Faculty, (5) International Students and (6) Citations per Faculty. Its 2019 edition data is on website: https://www.topuniversities.com/ university-rankings/world-university-rankings/2019. We only adopt the data of the first 100 universities, ignoring the weights of the items applied in the QS ranking, to

demonstrate the aggregation method proposed by this paper. For simplicity, we only use the ranking order in stead of the university name. The first six lines of the data is listed in Table 1. We scale the data by using Eq. (7), some of the scaled data, $\overline{S} = [\overline{s}_{ki}]_{100 \times 6}$, is shown in Table 2.

Table 1: Head of QS 2019 university ranking

Rank	Academic Reputation	Employer Reputation	Faculty Student	International Faculty	International Students	Citations per Faculty
1	100.0	100.0	100.0	100.0	95.5	99.8
2	100.0	100.0	100.0	99.8	70.5	99.0
3	100.0	100.0	99.3	92.1	75.7	99.8
4	98.7	81.2	100.0	96.8	90.3	100.0
5	100.0	100.0	100.0	99.6	98.8	83.0
6	100.0	100.0	100.0	99.4	97.9	77.2

Table 2: Some rows of scaled QS ranking data

	Academic Reputation	Employer Reputation	Faculty Student	International Faculty	International Students	Citations per Faculty
3	1.091440565	1.095582406	1.043613932	0.616368388	0.260404733	1.357432899
5	1.091440565	1.095582406	1.069509142	0.867988403	1.058619878	0.630398988
16	1.073787331	0.962765963	1.069509142	-1.118132246	0.968777481	-0.399565721
21	0.485346225	-1.106105555	1.069509142	0.207066497	0.464277865	0.509226669
34	0.208778905	-0.278556948	1.062110511	-0.057973251	-0.461789922	-0.001428102
46	-0.173707815	0.022834212	-0.939219245	0.153387561	1.100085600	1.167019256
70	-0.650345111	-0.968180787	-0.384321901	0.690176925	0.986054865	-0.486117377
79	-1.050485063	0.324225371	-0.646973311	0.596238787	0.692339336	-0.862617081
83	-2.515703419	-1.499446560	1.062110511	-0.910126368	-2.168795471	1.348777734
94	-0.844530676	0.268033799	-0.783847989	0.364748373	0.464277865	-1.156892712

5.1. PCA and its generated basic data

By using Eq. (11), we have the partial correlation coefficients of all items/columns, PC_{ij}, $i, j = 1, \ldots, 6$, see Table 3. Executing the PCA method, we have six new orthogonal variables $V_i = v_{i1}\overline{s}_1 + \cdots + v_{i6}\overline{s}_6$, v_{ij}, $i, j = 1, \ldots, 6$, and the corresponding eigenvalues, λ_i, $i = 1, \ldots, 6$, shown in Table 4. From Eqs. (9), (10) and (12), we get the partial importance, PI_i, comprehensive importance, CI_i, and comprehensive correlation coefficients, CC_{ij}, $i, j = 1, \ldots, 6$, as shown in Table 5.

Table 3: Partial correlation coefficients of six items

	Academic Reputation	Employer Reputation	Faculty Student	International Faculty	International Students	Citations per Faculty
Academic Reputation	1	0.638098883	0.033301470	-0.008442227	0.114413368	0.107028200
Employer Reputation	0.638098883	1	0.094292559	0.178510026	0.261781288	-0.020298971
Faculty Student	0.033301470	0.094292559	1	-0.071973643	-0.063081537	-0.261466578
International Faculty	-0.008442227	0.178510026	-0.071973643	1	0.669626119	0.177230965
International Students	0.114413368	0.261781288	-0.063081537	0.669626119	1	0.009998829
Citations per Faculty	0.107028200	-0.020298971	-0.261466578	0.177230965	0.009998829	1

Table 4: Coefficients of new orthonogonal variables and their eigenvalue

Items	V_1	V_2	V_3	V_4	V_5	V_6
Academic Reputation	0.579011568	0.623970958	-0.360404140	-0.030728344	0.307443796	-0.223715857
Employer Reputation	0.712207757	0.553769531	-0.087385473	-0.093531743	-0.370647994	0.179819560
Faculty Student	-0.068606417	0.483220825	0.605759311	0.626324906	0.045570206	0.022066242
International Faculty	0.701338267	-0.529131956	0.270493049	0.133131982	-0.179312649	-0.323890156
International Students	0.759926870	-0.365521491	0.353186259	-0.107125941	0.282542163	0.269923353
Citations per Faculty	0.205803332	-0.357301225	-0.707614168	0.560794304	0.000673998	0.121541380
Eigenvalues λ_i	1.958920350	1.470948908	1.203096737	0.745665649	0.345961821	0.275406535

Table 5: Comprehensive correlation coefficients and partial, comprehensive importance

	Academic Reputation	Employer Reputation	Faculty Student	International Faculty	International Students	Citations per Faculty	Partial importance	Comprehensive importance
Academic Reputation	1	0.582353302	-0.334808558	-0.020696863	-0.078293903	0.143601928	2.676557683	0.178851286
Employer Reputation	0.582353302	1	-0.242432118	0.180225465	0.103292379	0.006404549	2.562355078	0.171220110
Faculty Student	-0.334808558	-0.242432118	1	-0.129802039	-0.647491543	-0.296385353	2.062846409	0.137842250
International Faculty	-0.020696863	0.180225465	-0.129802039	1	0.750530051	0.179901256	2.728401824	0.182315583
International Students	-0.078293903	0.103292379	-0.647491543	0.750530051	1	0.051670322	2.703184476	0.180630525
Citations per Faculty	0.143601928	0.006404549	-0.296385353	0.179901256	0.051670322	1	2.231923994	0.149140248

5.2. Maximum entropy and compromise methods

For the comprehensive importance in the last column of Table 5, we have $CI_4 > CI_5 > CI_1 > CI_2 > CI_6 > CI_3$, which can be transformed into the following preference constraints by according to Eq. (21):

71

_____

$$I_\mu(\{4\}) - I_\mu(\{5\}) \geq 1/24,$$
$$I_\mu(\{5\}) - I_\mu(\{[1]\}) \geq 1/24,$$
$$\ldots \tag{21}$$
$$I_\mu(\{6\}) - I_\mu(\{3\}) \geq 1/24,$$

where $1/24 = \min(0.05, 1/(4 \times 6))$ is basically the threshold δ_1. Since
$$CC_{12}, CC_{16}, CC_{24}, CC_{25}, CC_{26}, CC_{45}, CC_{46}, CC_{56} > 0,$$
and
$$CC_{13}, CC_{14}, CC_{15}, CC_{23}, CC_{34}, CC_{35}, CC_{36} < 0,$$
from Eq. (15), we have
$$
\begin{aligned}
I_\mu(\{1,2\}), I_\mu(\{1,6\}), I_\mu(\{2,4\}), I_\mu(\{2,5\}) &\in [-1, -0.05], \\
I_\mu(\{2,6\}), I_\mu(\{4,5\}), I_\mu(\{4,6\}), I_\mu(\{5,6\}) &\in [-1, -0.05], \\
I_\mu(\{1,3\}), I_\mu(\{1,4\}), I_\mu(\{1,5\}), I_\mu(\{2,3\}) &\in [0.05,1], \\
I_\mu(\{3,4\}), I_\mu(\{3,5\}), I_\mu(\{3,6\}) &\in [0.05,1].
\end{aligned} \tag{22}
$$
Then according to maximum entropy principle model in Eq. (2), we have the following MGLP:
$$\min z = \sum_{A \subset N} d_A^+ + d_A^-$$
$$s.t. \begin{cases} \text{The boundary and monotonicity condition of capacity,} \\ \text{Eqs. (21) and (21)(22),} \\ \mu(A) - d_A^+ + d_A^- = \frac{|A|}{6}, d_A^+, d_A^- \geq 0, \forall A \subset \{1,2,3,4,5,6\}. \end{cases} \tag{23}$$
And according to compromise principle model in Eq. (3), we have
$$\min z = \sum_{k=1}^{100} d_k^+ + d_k^-$$
$$s.t. \begin{cases} \text{The boundary and monotonicity condition of capacity,} \\ \text{Eqs. (21) and (21)(22),} \\ C(x^k) - d_k^+ + d_k^- = \max x^k, d_k^+, d_k^- 0, k = 1,\ldots,100, \end{cases} \tag{24}$$

where x^k is k-th row of the scaled QS ranking data, e.g., see the first row of Table 2 $x^3 = (1.0914406, 1.0955824, 1.0436139, 0.6163684, 0.2604047, 1.3574329)$ and $\max x^3 = 1.3574329$. Solving the above two models, we have the two optimal capacities, whose interaction indices are listed in Table 6. One can see that the maximum entropy principle tends to keep the same cardinality subsets with similar index value and by comparison the compromise principle leads to a bit variety on the index values.

Table 6: The interaction indices obtained by maximum entropy and compromise methods

A	Max Entropy	Compromise	A	Max Entropy	Compromise	A	Max Entropy	Compromise	A	Max Entropy	Compromise
∅	0.5121	0.5735	{4,5}	-0.0500	-0.1756	{1,2,6}	0.0780	0.1335	{1,2,4,6}	-0.1098	-0.2175
{1}	0.1875	0.1875	{1,6}	-0.0500	-0.0500	{1,3,6}	0.0050	-0.0289	{1,3,4,6}	0.1219	0.0881
{2}	0.1458	0.1458	{2,6}	-0.0500	-0.0786	{2,3,6}	0.0422	0.0230	{2,3,4,6}	0.0255	-0.0674
{3}	0.0625	0.0625	{3,6}	0.0500	0.0500	{1,4,6}	0.1170	0.0326	{1,2,5,6}	-0.0078	-0.2646
{4}	0.2708	0.2708	{4,6}	-0.0500	-0.0953	{2,4,6}	0.0631	0.0391	{1,3,5,6}	0.1284	0.0881
{5}	0.2292	0.2292	{5,6}	-0.0500	-0.1322	{3,4,6}	0.0314	-0.0398	{2,3,5,6}	-0.0255	-0.0674
{6}	0.1042	0.1042	{1,2,3}	0.0174	0.0330	{1,5,6}	0.0772	0.0984	{1,4,5,6}	-0.0646	0.0197
{1,2}	-0.0500	-0.0500	{1,2,4}	0.1596	0.1893	{2,5,6}	0.0845	0.0627	{2,4,5,6}	-0.0412	-0.0716
{1,3}	0.0500	0.0500	{1,3,4}	0.1031	0.1164	{3,5,6}	0.0305	-0.0398	{3,4,5,6}	-0.0317	0.1144
{2,3}	0.0500	0.0500	{2,3,4}	0.0637	0.0483	{4,5,6}	0.1568	0.2344	{1,2,3,4,5}	0.3181	0.0848
{1,4}	0.0500	0.0500	{1,2,5}	0.1269	0.0654	{1,2,3,4}	0.1027	0.1231	{1,2,3,4,6}	0.1128	0.0678
{2,4}	-0.0500	-0.1080	{1,3,5}	0.1245	0.1164	{1,2,3,5}	0.1472	0.1231	{1,2,3,5,6}	0.1371	0.0678
{3,4}	0.0500	0.0500	{2,3,5}	0.0431	0.0483	{1,2,4,5}	-0.0941	0.1598	{1,2,4,5,6}	0.6593	0.3156
{1,5}	0.0500	0.0500	{1,4,5}	0.1713	0.0461	{1,3,4,5}	0.1411	0.0388	{1,3,4,5,6}	0.0476	-0.1008
{2,5}	-0.0500	-0.0500	{2,4,5}	0.1553	0.2180	{2,3,4,5}	0.1406	0.0968	{2,3,4,5,6}	0.0733	0.0153
{3,5}	0.0500	0.0500	{3,4,5}	0.0518	0.1251	{1,2,3,6}	-0.1412	-0.1068	{1,2,3,4,5,6}	0.2505	-0.2809

_____

5.3. MCCPI method

From the partial importance listed in last second column in Table 5, the partial correlation coefficients listed in Table 3 and Eq. (16), we can have

$$I_1^{12} = \frac{PI_1}{PI_1+PI_2} = 0.5108995,$$
$$\cdots$$
$$I_5^{56} = \frac{PI_5}{PI_5+PI_6} = 0.5023214,$$
$$I_{12}^{P} = -PC_{12} = -0.638098883,$$
$$\cdots$$
$$I_{56}^{P} = -PC_{56} = -0.009998829.$$

(25)

From model 6, we have

$$\min z = \sum_{i,j\in\{1,2,3,4,5,6\},i<j} d_{ij}^+ + d_{ij}^- + e_{ij}^+ + e_{ij}^-$$

$$s.t. \begin{cases} \text{The boundary and monotonicity condition of capacity,} \\ 0.5108995(I_\mu(\{1\}) + I_\mu(\{2\})) - I_\mu(1) - d_{12}^+ + d_{12}^- = 0, \\ \cdots \\ 0.5023214(I_\mu(\{5\}) + I_\mu(\{6\})) - I_\mu(5) - d_{56}^+ + d_{56}^- = 0, \\ -0.638098883(I_\mu(\{1\}) + I_\mu(\{2\})) - I_\mu(\{1,2\}) - e_{12}^+ + e_{12}^- = 0, \\ \cdots \\ -0.009998829(I_\mu(\{5\}) + I_\mu(\{6\})) - I_\mu(\{5,6\}) - e_{56}^+ + e_{56}^- = 0, \end{cases}$$

(26)

where $d_{ij}^+, d_{ij}^-, e_{ij}^+, e_{ij}^- \leq 0$, $i,j \in \{1,2,3,4,5,6\}, i < j$. Solving this MGLP, we get the optimal capacity and its Shapely interaction index, denoted as I_1^m. Furthermore, based on the Eq. (17), we set the $\delta_3 = 0.05, \delta_4 = 0.1$, execute Algorithm 1 with $\mathcal{K} = 10$, and denote the optimal Shapley interaction indices as I_2^m, ..., I_{11}^m. The table 7 shows these indices of singletons and pairs items, where I_1^m can be seen as the center surrounding by the random generation MCCPI generated indices, $I_2^m \sim I_{11}^m$. All these indices allow us to see more overall aggregation results or different dominance situations among the instances.

Table 7: The interaction indices obtained by MCCPI methods

	I_1^m	I_2^m	I_3^m	I_4^m	I_5^m	I_6^m	I_7^m	I_8^m	I_9^m	I_{10}^m	I_{11}^m
∅	0.6126	0.6077	0.6083	0.6307	0.6146	0.5802	0.6096	0.6203	0.6200	0.5837	0.6031
{1}	0.1789	0.1671	0.1880	0.1947	0.1628	0.1735	0.1623	0.1852	0.1964	0.1894	0.1839
{2}	0.1712	0.1784	0.1675	0.1628	0.1787	0.1667	0.1707	0.1845	0.1693	0.1695	0.1826
{3}	0.1378	0.1361	0.1421	0.1373	0.1369	0.1373	0.1250	0.1420	0.1359	0.1435	0.1344
{4}	0.1823	0.1685	0.1654	0.1871	0.1890	0.1878	0.2038	0.1690	0.1793	0.1708	0.1762
{5}	0.1806	0.1889	0.1861	0.1674	0.1881	0.1815	0.1961	0.1713	0.1788	0.1742	0.1622
{6}	0.1491	0.1610	0.1509	0.1507	0.1446	0.1533	0.1421	0.1480	0.1403	0.1526	0.1607
{1,2}	-0.2234	-0.2527	-0.2393	-0.2324	-0.2508	-0.2068	-0.1825	-0.2707	-0.2205	-0.2647	-0.2207
{1,3}	-0.0105	0.0106	-0.0409	-0.0103	-0.0120	-0.0135	-0.0315	-0.0104	0.0014	-0.0366	-0.0396
{2,3}	-0.0291	-0.0530	-0.0376	-0.0232	-0.0156	-0.0171	-0.0545	-0.0178	-0.0029	-0.0110	-0.0234
{1,4}	0.0030	0.0029	0.0195	0.0044	-0.0186	0.0365	-0.0222	0.0178	-0.0294	0.0037	0.0255
{2,4}	-0.0631	-0.0426	-0.0862	-0.0797	-0.0439	-0.0450	-0.1011	-0.0730	-0.0285	-0.0646	-0.0862
{3,4}	0.0230	0.0459	0.0246	0.0329	0.0408	0.0373	0.0102	0.0196	0.0112	0.0124	0.0174
{1,5}	-0.0411	-0.0679	-0.0770	-0.0651	-0.0667	-0.0333	-0.0247	-0.0355	-0.0599	-0.0325	-0.0417
{2,5}	-0.0921	-0.1213	-0.1043	-0.1016	-0.1035	-0.1211	-0.0833	-0.1278	-0.0910	-0.0583	-0.1202
{3,5}	0.0201	0.0344	0.0098	-0.0046	0.0527	0.0118	0.0348	0.0111	0.0119	0.0010	0.0096
{4,5}	-0.2430	-0.2187	-0.2229	-0.2678	-0.2430	-0.2222	-0.2498	-0.2303	-0.2562	-0.2548	-0.1980
{1,6}	-0.0351	-0.0581	-0.0628	-0.0630	-0.0486	-0.0225	-0.0502	-0.0130	-0.0530	-0.0352	-0.0549
{2,6}	0.0065	0.0347	0.0170	-0.0133	0.0035	0.0369	0.0288	0.0039	0.0101	0.0243	-0.0266
{3,6}	0.0750	0.0491	0.0675	0.0699	0.0742	0.0987	0.0816	0.0538	0.0483	0.0729	0.1044
{4,6}	-0.0587	-0.0358	-0.0515	-0.0515	-0.0831	-0.0288	-0.0616	-0.0741	-0.0418	-0.0530	-0.0612
{5,6}	-0.0033	-0.0099	-0.0107	-0.0254	-0.0061	0.0014	-0.0118	-0.0315	-0.0258	0.0174	0.0042

5.4. Comprehensive information oriented method

In order to use the comprehensive importance and correlation coefficients generated by PCA to obtain some competent capacities, see Table 5, we can adopt Eq. (18) and the model in Eq. (19) to construct the following MGLP:

73

Jian-Zhang Wu, Yan-Qing Li, Feng-Feng Chen, Gleb Beliakov

_____

$$\min z = \sum_{A \subset \{1,2,3,4,5,6\}, |A|=1,2} d_A^+ + d_A^-$$

$$s.t. \begin{cases} \text{The boundary and monotonicity condition of capacity,} \\ I_\mu(\{1\}) - 0.1788513 - d_1^+ + d_1^- = 0, d_1^+, d_1^- = 0 \geq 0, \\ \quad \dots \\ I_\mu(\{6\}) - 0.1491402 - d_2^+ + d_2^- = 0, d_2^+, d_2^- = 0 \geq 0, \\ I_\mu(\{1,2\}) + 0.582353302 - d_{12}^+ + d_{12}^- = 0, d_{12}^+, d_{12}^- \geq 0, \\ \quad \dots \\ I_\mu(\{5,6\}) + 0.051670322 - d_{56}^+ + d_{56}^- = 0, d_{56}^+, d_{56}^- \geq 0. \end{cases} \quad (27)$$

Solving the above model, we achieve the optimal capacity and its Shapley interaction index, denoted as I_1^c. In order to observe more neighboring situations, we execute Algorithm 2 with $\delta_6 = 0.05$, $\delta_4 = 0.1$, and $\mathcal{K} = 10$, and denote the optimal Shapley interaction indices as I_2^m, ..., I_{11}^m, see Table 8 which bears kind of likeness to Table 7.

Table 8: The interaction indices obtained by comprehensive information oriented methods

	I_1^c	I_2^c	I_3^c	I_4^c	I_5^c	I_6^c	I_7^c	I_8^c	I_9^c	I_{10}^c	I_{11}^c
∅	0.5670	0.5716	0.4810	0.5717	0.5420	0.4819	0.5269	0.6210	0.5396	0.4801	0.5725
{1}	0.1789	0.1562	0.2107	0.2121	0.2106	0.1378	0.2060	0.1499	0.1990	0.1759	0.1730
{2}	0.1712	0.1844	0.1361	0.2188	0.1518	0.1605	0.2059	0.1836	0.1509	0.1455	0.1884
{3}	0.1378	0.1582	0.1823	0.1342	0.1563	0.2313	0.1606	0.1016	0.1434	0.1784	0.1332
{4}	0.1823	0.2230	0.1474	0.1415	0.1547	0.1467	0.1769	0.2248	0.1504	0.2103	0.2120
{5}	0.1806	0.1431	0.1931	0.1457	0.1703	0.2168	0.1379	0.1826	0.1648	0.1825	0.1121
{6}	0.1491	0.1350	0.1304	0.1478	0.1563	0.1070	0.1127	0.1574	0.1915	0.1075	0.1814
{1,2}	-0.2232	-0.1894	-0.1566	-0.3712	-0.2528	-0.3075	-0.3401	-0.1317	-0.2590	-0.1314	-0.2831
{1,3}	0.1043	0.1150	0.1651	0.0456	0.0901	0.1169	0.0707	0.0825	0.0629	0.1423	0.0607
{2,3}	0.0311	0.0544	0.0502	0.0611	0.0675	0.0791	0.0282	0.0625	0.0672	0.0481
{1,4}	0.0207	-0.0416	0.0969	-0.0092	-0.0020	-0.0021	0.0829	0.0139	0.0929	0.1168	0.0377
{2,4}	-0.1802	-0.2230	-0.0802	-0.0976	-0.1008	-0.0909	-0.1400	-0.2023	-0.1586	-0.1931	-0.1973
{3,4}	0.0666	0.1061	0.0721	0.0611	0.0554	0.0958	0.1017	0.0515	0.0726	0.1162	0.0612
{1,5}	0.0783	0.1107	0.1229	0.0707	0.0767	0.1072	0.0828	0.0173	0.0033	0.1537	0.1149
{2,5}	-0.1033	-0.0600	-0.1785	-0.1326	-0.1433	-0.0165	-0.0796	-0.1938	-0.0812	-0.0726	-0.0057
{3,5}	0.0945	0.0973	0.1365	0.0968	0.1248	0.1701	0.0689	0.0306	0.0742	0.0985	0.0636
{4,5}	-0.2173	-0.1500	-0.2227	-0.1647	-0.1531	-0.2115	-0.2512	-0.3292	-0.2345	-0.2674	-0.1999
{1,6}	-0.1436	-0.0557	-0.1233	-0.1098	-0.2341	-0.1199	-0.2079	-0.1585	-0.2190	-0.0662	-0.1974
{2,6}	-0.0064	-0.1054	0.0501	0.0324	0.0740	0.0171	0.0787	-0.0990	0.0875	0.0170	-0.0082
{3,6}	0.0706	0.0333	0.0823	0.0741	0.0772	0.1022	0.0791	0.0664	0.1064	0.0672	0.0862
{4,6}	-0.1799	-0.2605	-0.1810	-0.1689	-0.2111	-0.1008	-0.0984	-0.1428	-0.1212	-0.1304	-0.1760
{5,6}	-0.0517	-0.0010	-0.0665	-0.0967	-0.0099	-0.0829	-0.0640	0.0278	-0.1339	-0.0997	-0.1144

One may concern the final aggregation result where different capacities may more or less lead to different situations. Table 9 shows the ranking orders obtained by the four types of capacity identifications, in the title row, QS stands for the QS ranking order, R1 the maximum entropy method, R2 the compromise method, R3 the MCCPI method and R4 the comprehensive information oriented method. The ranking order of R3 is just the average of the 11 ranking orders achieved with their 11 capacities, so for R4. The difference between the QS ranking order and other four orders is that QS applies weights of 0.4, 0.1, 0.2, 0.05, 0.05, 0.2 to the six items in sequence, where the other four orders adopt a relatively even weight, see the Shapley importance in Tables 6, 7 and 8. One can figure out that R1 and R2 bear good likeness meanwhile R3 and R4 have the relatively similar result. The similarity and difference of the aggregation results mainly lie in the applied principles in the capacity identification and the involved parameters also put some effects on the final outcomes.

_____

Table 9: The final ranks of four types of methods

QS	R1	R2	R3	R4	QS	R1	R2	R3	R4	QS	R1	R2	R3	R4	QS	R1	R2	R3	R4
1	1	1	1	1	26	82	80	52	43	51	43	46	50	37	76	53	50	60	55
2	9	9	2	4	27	29	27	25	40	52	31	36	44	39	77	100	99	99	99
3	8	7	4	8	28	17	17	25	18	53	79	87	77	76	78	65	65	70	71
4	7	8	5	5	29	14	13	17	15	54	25	23	31	34	79	48	51	63	58
5	2	2	3	2	30	39	39	39	26	55	52	54	38	73	80	80	77	84	73
6	3	4	6	6	31	16	14	23	16	56	63	53	61	65	81	74	72	82	89
7	4	6	7	9	32	45	31	46	42	57	67	66	66	54	82	57	57	65	60
8	5	3	8	3	33	18	19	26	21	58	90	91	89	91	83	96	88	93	90
9	22	21	13	14	34	58	68	57	51	59	42	45	48	67	84	99	100	98	100
10	6	5	9	7	35	85	86	73	82	60	71	71	49	52	85	54	58	68	75
11	10	11	10	11	36	84	82	69	84	61	51	47	59	36	86	95	96	92	88
12	12	15	12	13	37	26	26	36	31	62	72	78	78	63	87	78	84	64	96
13	19	16	18	17	38	23	24	21	32	63	88	81	87	87	88	89	92	95	93
14	15	18	15	19	39	27	32	27	44	64	87	93	90	82	89	75	83	88	78
15	24	25	22	21	40	81	73	79	70	65	59	62	33	46	90	92	90	91	95
16	37	29	20	23	41	50	52	51	38	66	73	75	85	75	91	49	55	72	80
17	55	48	32	35	42	28	33	30	53	67	97	98	97	98	92	64	64	81	62
18	11	10	14	12	43	60	49	54	56	68	62	63	58	47	93	76	79	75	68
19	35	41	42	29	44	46	43	35	33	69	70	67	80	66	94	56	61	71	64
20	36	37	29	28	45	33	40	34	59	70	44	35	55	45	95	93	94	96	92
21	40	44	28	25	46	32	34	37	41	71	83	76	76	83	96	69	70	67	77
22	21	20	11	22	47	38	42	45	48	72	94	95	94	94	97	68	69	54	86
23	77	74	56	61	48	34	38	43	57	73	66	59	74	49	98	98	97	100	97
24	20	22	19	30	49	30	30	41	24	74	47	60	47	50	99	91	85	62	85
25	13	12	16	10	50	41	28	40	27	75	86	89	86	79	100	61	56	83	69

6. Conclusions

In this paper, we combined the PCA method and capacity theory to construct some self-organized and unsupervised multiple criteria/items information aggregation models, where the PCA provides the importance and correlation information as the input preference information to guide the Choquet capacity and integral aggregation scheme.

The maximum entropy principle tends to ensure the most even contributions of all the items and the compromise principle tries to provide the most equal chance for all the instance to reach their excellent performances. Both two principles have a trend to close the symmetric additive capacity. The MCCPI and the comprehensive information oriented methods aim to minimize the deviation between the competent capacities and the partial and comprehensive importance and interaction derived from the PCA. And the random generation and simulation algorithms can be applied for these two methods to explore the neighbor situation and provide a more comprehensive view of the aggregation outcomes.

The combination of different aggregation methods is also good way to obtain more useful analysis supports and decision aids. It should be mentioned that in all the models proposed in this paper, additional explicit or implicit supervised preference, especially about the 3 or higher order subsets, as well as the inconsistency check and adjustment algorithms can be added and executed simultaneously.

Future research will focus on some real applications of multiple criteria analyses, such as the classification results fusion and recommendation systems integration. And finally we want to emphasis that the complexity inherent in the structure of capacity actually provides many potential chances to store more knowledge and to collaborate with some artificial intelligence algorithms more flexibly.

Jian-Zhang Wu, Yan-Qing Li, Feng-Feng Chen, Gleb Beliakov

_____

REFERENCES

[1]**G. Beliakov, S. James and J.-Z. Wu (2019),** *Discrete Fuzzy Measures: Computational Aspects. Springer,* Cham, Switzerland;

[2] **G. Beliakov and J.-Z. Wu(2019),** *Learning Fuzzy Measures from Data: Simplifications and Optimisation Strategies. Information Sciences,* 494:100–113;

[3] **G. Beliakov(2009),** *Construction of Aggregation Functions from Data Using Linear Programming. Fuzzy Sets and Systems*, 160(1):65–75;

[4] **G. Choquet (1954),** *Theory of Capacities. Annales de l'institut Fourier,* 5:131–295;

[5] **A. Fallah Tehrani, W. Cheng, K. Dembczynski and E. Hullermeier** *(2012), Learning Monotone Nonlinear Models Using the Choquet Integral. Machine Learning,* 89(1-2):183–211, 2012;

[6] **K. Fujimoto, I. Kojadinovic, and J.-L. Marichal (2006),** *Axiomatic Characterizations of Probabilistic and Cardinal-Probabilistic Interaction Indices. Games and Economic Behavior,* 55(1):72–99;

[7] **M. Grabisch, I. Kojadinovic and P. Meyer (2008),** *A Review of Methods for Capacity Identification in Choquet Integral Based Multi-Attribute Utility Theory: Applications of the Kappalab R Package. European Journal of Operations Research,* 186(2):766–785, 2008;

[8] M. Grabisch. k-order additive discrete fuzzy measures and their representation. Fuzzy Sets and Systems, 92(2):167–189, 1997.

[9] **M. Grabisch (2016),** *Set Functions, Games and Capacities in Decision Making. Springer,* Berlin, New York;

[10] **I. T. Jolliffe and J. Cadima (2016),** *Principal Component Analysis: A Review and Recent Developments. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences,* 374(2065):20150202;

[11] **A. Kassambara(2017),** *Practical Guide to Principal Component Methods in R: PCA, M (CA), FAMD, MFA,* HCPC, factoextra, volume 2. STHDA;

[12] **I. Kojadinovic(2007),** *Minimum Variance Capacity Identification. European Journal of Operational Research,* 177(2):498–514;

[13] **J.-L. Marichal (2002),** *Entropy of Discrete Choquet Capacities. European Journal of Operations Research*, 137(3):612–624;

[14] **P. Meyer and M. Roubens (2005),** *Choice, Ranking and Sorting in Fuzzy Multiple Criteria Decision Aid.* In J. Figueira, S. Greco, and M. Ehrogott, editors, Multiple Criteria Decision Analysis: State of the Art Surveys, pages 471–503. *Springer New York;*

_____

[15] **M. Sugeno (1974),** ***Theory of Fuzzy Integrals and Its Applications.*** PhD thesis, Tokyo Institute of Technology;

[16] **Z. Wang and G. J. Klir (2010),** *Generalized Measure Theory*. *Springer Science & Business Media*, New York;

[17] **S. J. Wetzel (2017),** ***Unsupervised Learning of Phase Transitions: From Principal Component Analysis to Variational Autoencoders.*** *Physical Review E,* 96(2):022140;

[18] J.-Z. Wu and G. Beliakov. Nonadditivity index and capacity identification method in the context of multicriteria decision making. Information Sciences, 467:398–406, 2018.

[19] **J.-Z. Wu and G. Beliakov(2019),** ***Nonmodularity Index for Capacity Identifying with Multiple Criteria Preference Information.*** *Information Sciences,* 492:164–180;

[20] **J.-Z. Wu and G. Beliakov (2019),** ***Probabilistic Bipartition Interaction Index of Multiple Decision Criteria Associated with the Nonadditivity of Fuzzy Measures.*** *International Journal of Intelligent Systems,* 34(2):247–270;

[21] **J.-Z. Wu, L. Huang, R.-J. Xi and Y.-P. Zhou (2019), Multiple *Goal Linear Programming-Based Decision Preference Inconsistency Recognition and Adjustment Strategies.*** *Information,* 10(7):223;

[22] J**.-Z. Wu, S. Yang, Q. Zhang, and S. Ding (2015),** ***2-Additive Capacity Identification Methods from Multicriteria Correlation Preference Information.*** *IEEE Transactions on Fuzzy Systems,* 23(6):2094–2106;

[23] **J.-Z. Wu, Q. Zhang, Q. Du and Z. Dong (2014),** ***Compromise Principle Based Methods of Identifying Capacities in the Framework of Multicriteria Decision Analysis.*** *Fuzzy Sets and Systems,* 246:91–106;

[24] **J.-Z. Wu, Y.-P. Zhou, L. Huang, and J.-J. Dong (2019),** ***Multicriteria Correlation Preference Information (MCCPI) Based Ordinary Capacity Identification Method.*** *Mathematics,* 7(3):300, 2019.