博士学位論文

On a Fuzzification and Optimization Problems of Clustering Indices

近畿大学大学院

産業理工学研究科産業理工学専攻

Devi Rahmah Sope

Devi Rahmah Sope

博士学位論文

On a Fuzzification and Optimization Problems of Clustering Indices

令和 4 年 10 月 12 日

近畿大学大学院

産業理工学研究科産業理工学専攻

Devi Rahmah Sope

Acknowledgement

Praise God Almighty, for the presence of plenty of mercy and His grace, so that the writer can complete the thesis with the title: On a Fuzzification and Optimization Problems of Clustering Indices. This is to meet one of the conditions and completed studies in order to obtained a Doctor of Engineering, Deparment of Informatics and Computer Sciences, Faculty of Humanity-Oriented Sciences and engineering, Kindai University, Japan. Appreciation and thanks to Prof. Fujio Mitsuhiko, who always support me and teach me about many things especially Mathematics.

Appreciation and thanks to Beasiswa Unggulan who chose me as a scholarship awardee, so I can pursue my dream to continue to study abroad. Thank you very much for my beloved Father, Drs. La Imba, my beloved Mother Wa Ode Eny Sabaniah and my Sisters/Brothers who have put all our love and affection and attention to moral and willing to take care of my Son (Naoki-Kun) in Indonesia while I'm in Japan. Thans so much for my Husband Laode Sulaiman who always supports me to gain my degree. May Allah SWT always bestow grace, health, gifts and blessings in the world and in the hereafter over budi good that has been given to the author.

International Journal of Mathematics for Industry
Vol. 13, No. 1 (2023) 1–30

Original Article

On a fuzzification and optimization problems of clustering indices

Devi Rahmah Sope[†,§] and Mitsuhiko Fujio[†,‡]
[†]*Department of Informatics and Computer Sciences, Kindai University, Iizuka City, 820-8555, Japan*
[§]*2044954001v@ed.fuk.kindai.ac.jp*
[‡]*fujio@fuk.kindai.ac.jp*

February 23, 2023

Results of clustering are qualitatively evaluated by quantities called clustering indices. While many clustering indices are proposed, in Ref. 2, Desgraupes reviewed 27 indices, most of which are applicable only to crisp clustering and only one of which is applicable to fuzzy clustering. In the previous article Ref. 1, the authors gazed on analytic 3 indices among them and modified them to fit fuzzy clustering by regarding the membership degrees as distribution functions of objects over clusters. In this article, this method called a fuzzification, is applied to all the indices of Desgraupes. This investigation also includes optimization of indices. A significant benefit of fuzzy clustering is that the membership degrees allow the optimization problems to be treated as continuous, whereas the ones in the crisp case are discrete, making it generally easier to solve. After giving a precise description of fuzzification which is briefly given in Ref. 1, the authors fuzzify all of 27 indices of Desgraupes and then solve the fuzzy clustering problems having 13 analytic indices among them as the objective functions by the gradient method.

Keywords: Clustering indices, fuzzy c means, fuzzification, optimization.

1. Introduction

Clustering is an unsupervised learning method that groups data into clusters based on their characteristics and similarities. The most widely used types of clustering are crisp clustering and fuzzy clustering. K-Means is one of the most popular crisp clustering methods which attempt to minimize an objective function and as a result, each data belongs exclusively to one cluster.[19] On the other hand, in fuzzy clusterings, Fuzzy C-Means (FCM) is known as one of the most popular methods, where each data is allowed to belong to one or more clusters determined by a quantity μ called the membership degree.[3] FCM algorithm is processed by optimizing an objective function containing the membership degree with an exponent $m > 1$. The exponent m is called the fuzziness.[5] The behaviors of optimal clustering in the limit transitions of m, not only for $m \to 1$ but also for $m \to \infty$, are mentioned in Ref. 4. The observation there tells that crisp clusterings are achieved as special cases of fuzzy clusterings through the membership degree μ and the fuzziness m. In Ref. 1, we proposed a method to enable crisp indices to be applicable to fuzzy clustering. We call the method a fuzzification.

In general, fuzzification is a process of transforming a crisp set to a fuzzy set or a process of converting a crisp quantity into a fuzzy quantity. Some authors mentioned about the fuzzification in a procedural sense. Namely, in Ref. 9, Nolan researched on a prototype application of fuzzy logic and expert systems in education assessment. In his article, a fuzzification done, rather than fuzzifying the expression of the indicator, by giving a method in the fuzzy expert system to convert the crisp input into

linguistic value such as high (insightful, thoughtful), medium (basic, nofrills) or low (confused, missing pieces). And he also used the trapezoidal fuzzifier to generate the membership degree. On the other hand, in Ref. 12, Parkhira et al., researched a validity index (PBM Index) and its fuzzification and compare it with the other indices for crisp such as *Davies Bouldin Index* and *Dunn's Index* and one validiy index for fuzzy clustering namely *Xie Beni Index*. They proposed the fuzzification by using the fuzziness $m = 1.5$. However the fuzzification done only for 1 index namely PBM Index. In our article, we expand the research to fuzzify 27 crisp clustering indices which are consist of 13 analytic indices and 14 non-analytic indices. The 13 analytic indices are Ball-Hall Index, Banfeld-Raftery Index, Calinski-Harabasz Index, Det Ratio Index, Ksq DetW Index, Log Det Ratio Index, Log SS Ratio Index, McClain-Rao Index, Point-Biserial Index, Ratkowsky-Lance Index, Scott-Symons Index, Trace W Index, Trace WiB Index. And the 14 non-analytic indices are C Index, Davies-Bouldin Index, Dunn Index, Baker-Hubert Gamma Index, GDI Index, G plus Index, PBM Index, Ray-Turi Index, SD Index, S_Dbw Index, Silhouette index, Tau index, Wemmert-Gançarsky Index, Xie-Beni Index.

Clustering indices evaluate the quality of clustering results by measuring their compactness and separateness. The compactness is measured by the extent of objects within a cluster and, on the other hand, the separateness is measured by the distance between different clusters. The smaller the compactness and the greater the separateness, the better the clustering. In Ref. 2, Bernard Desgroupes has reviewed 27 clustering indices. These indices are originally designed for crisp clustering and only one (Xie Beni Index) is used in fuzzy clustering. In this article, we convert them applicable to fuzzy clustering by using the method of fuzzification which we have proposed in the previous work Ref. 1. And we restrict our attention to the analytic ones among them to be optimized as the new objective functions in fuzzy clustering. We approach this problem using the traditional method of conditional optimization problem by applying the gradient method to the objective function.

2. Clustering and Clustering Indices

2.1. *Clusterings*

We consider the clustering problem of given n data $o_i (1 \leq i \leq n)$ in R^d. Although treated as row vectors in Desgraupes' article, we denote data vectors in R^n as column vectors for convenience. In K-Means, the data are classified into c clusters, where each data belongs to exclusively to exactly one cluster $C_k (1 \leq k \leq c)$. On the other hand, FCM permits each data belongs to one or more clusters. This situation is represented by the quantity μ_{ik} called the membership degree which measures the assignment ratio of the i-th data to the k-th cluster. The membership degrees μ_{ik}'s satisfy the following conditions:

$$0 \leq \mu_{ik} \leq 1 \qquad (1 \leq i \leq n, 1 \leq k \leq c), \tag{1}$$

$$\sum_{k=1}^{c} \mu_{ik} = 1 \qquad (1 \leq i \leq n), \tag{2}$$

$$0 < \sum_{i=1}^{n} \mu_{ik} < n \qquad (1 \leq k \leq c). \tag{3}$$

The condition (1) comes from that μ_{ik} gives the assignment ratio. The condition (2) means that for each fixed i, the i-th data is distributed to clusters and summing up the μ_{ik} over k is equal to 1. The condition (3) says that for each fixed k, the clustering is not trivial i.e., each cluster is neither empty nor whole. To handle the membership degree systematically, we introduce the fuzzy matrix:

$$(\mu_{ik}) = \begin{pmatrix} \mu_{11} & \cdots & \mu_{1c} \\ \vdots & \ddots & \vdots \\ \mu_{n1} & \cdots & \mu_{nc} \end{pmatrix}. \tag{4}$$

This framework includes crisp clustering. In fact, for crisp cases, the constraint (1) is reduced to

$$\mu_{ik} \in \{0, 1\} \qquad (1 \leq i \leq n, 1 \leq k \leq c). \tag{5}$$

The membership degree μ_{ik} takes the value 1 if and only if the data o_i belongs to the cluster C_k otherwise it takes the value 0. This condition can be written as:

$$\mu_{ik} = \begin{cases} 0 & o_i \notin C_k \\ 1 & o_i \in C_k \end{cases}. \tag{6}$$

For crisp case we put

$$I_k = \{i | o_i \in C_k\}. \tag{7}$$

Then we have

$$\sum_{i=1}^{n} \mu_{ik} = n_k, \tag{8}$$

where n_k is the cardinality of the set I_k. Furthermore, by using I_k, the condition (3) can be rewritten as:

$$0 < \sum_{i \in I_k} \mu_{ik} < n \qquad (1 \leq k \leq c). \tag{9}$$

In both of FCM and K-Means, the objective function

$$J = \sum_{k=1}^{c} \sum_{i=1}^{n} \mu_{ik}^{m} \; d(o_i, z^{\{k\}})^2 \tag{10}$$

is minimized. Here, $d(o_i, z^{\{k\}})^2$ is the square distance between the i-th data o_i and the center $z^{\{k\}}$ of the k-th cluster. The objective function employs the membership degrees μ_{ik} not as it is but with an exponent $m > 1$. This exponent m is called the fuzziness which affects on the result of clustering. As m tends to infinity, μ_{ik}^m for each data tends to 0 unless $\mu_{ik} = 1$.[4] In Ref. 5, Pall and Bezdek suggested taking $m \in [1.5, 2.5]$ is suitable for FCM. Similar recommendations found in Refs. 6, 7, 8, 10. Moreover, by taking the limit $m \to 1$ the FCM will reduce to crisp clustering.[4]

2.2. *Clustering Indices*

Clustering algorithms produce optimal clusterings through clustering indices. A clustering index works by evaluating the qualities of clustering such as the compactness and the separateness of clusters. In this subsection, we will give a brief description about 27 crisp clustering indices which is mentioned as Internal Indices in Ref. 2. These indices consist of analytic indices and non-analytic indices. An analytic index means an index expressed by an analytic function, and in particular, all clustering indices of Desgraupes that do not include any statistical functions such as minimum or maximum function. On the other hand, a non-analytic index is one that includes at least one statistical function.

Before describing indices, we introduce key quantities commonly used in those clustering indices. These quantities, with the exception of slight changes in notation and symbols, all were brought from Desgraupes' article.

(a) The Scatter Matrix T is defined as the dispersion of data o_i with respect to the total center z:

$$T = \sum_{i=1}^{n} (o_i - z) \; {}^t(o_i - z). \tag{11}$$

Note that this quantity coincides with the variance-covariance matrix multiplied by the number of data n.

(b) The Total Scattering TSS is the trace of the matrix T which is equal to the sum of the squared distances (scattering) of the data around the total center:

$$TSS = \mathrm{Tr}(T) = \sum_{i=1}^{n} \big\|(o_i - z)\big\|^2. \tag{12}$$

(c) The Within Group Scatter Matrix $WG^{\{k\}}$ is the clusterwize dispersions of data in C_k with respect to the cluster center $z^{\{k\}}$:

$$WG^{\{k\}} = \sum_{i \in I_k} (o_i - z^{\{k\}}) \; {}^t(o_i - z^{\{k\}}). \tag{13}$$

Similarly to (11), this matrix is the variance-covariance matrix of data in C_k multiplied by the number n_k of data in it. And their sum for all the clusters is denoted by WG:

$$WG = \sum_{k=1}^{c} WG^{\{k\}}. \tag{14}$$

(d) The Within Group Scatter Dispersion $WGSS^{\{k\}}$ is the trace of matrix $WG^{\{k\}}$ and is equal to the sum of the scattering of the data in the cluster C_k around its center:

$$\begin{aligned} WGSS^{\{k\}} &= \mathrm{Tr}(WG^{\{k\}}) \\ &= \sum_{i \in I_k} \big\|(o_i - z^{\{k\}})\big\|^2. \end{aligned} \tag{15}$$

Also the sum of the within group scatter dispersions $WGSS$ is given as:

$$WGSS = \sum_{k=1}^{c} WGSS^{\{k\}}. \tag{16}$$

(e) The Between Group Scatter Matrix BG is the dispersion of the cluster centers $z^{\{k\}}$ with respect to the total center z, with the weight n_k:

$$BG = \sum_{k=1}^{c} n_k (z^{\{k\}} - z) \; {}^t(z^{\{k\}} - z). \tag{17}$$

(f) The Between Group Scatter Dispersion $BGSS$ is the trace of matrix BG which is equal to the weighted sum of the scattering of the cluster centers around the total center:

$$BGSS = \mathrm{Tr}(BG) = \sum_{k=1}^{c} n_k \big\|z^{\{k\}} - z\big\|^2. \tag{18}$$

(g) The total number of pairs of distinct data in the same clusters is defined by:

$$N_W = \sum_{k=1}^{c} \frac{n_k(n_k-1)}{2}. \tag{19}$$

Here we note that the k-th term $\frac{n_k(n_k-1)}{2}$ counts the number of pairs of distinct points in the cluster C_k. On the other hand, the number of pairs of data those do not belong to a same cluster is defined by:

$$N_B = \sum_{k<\ell} n_k n_\ell. \tag{20}$$

Then the total number of pairs of distinct data is:

$$N_T = \frac{n(n-1)}{2} = N_W + N_B. \tag{21}$$

Moreover, the sum of the within cluster distances is defined by:

$$\begin{aligned} S_W &= \sum_{k=1}^{c} \sum_{i,j\in I_k, i<j} \|o_i - o_j\| \\ &= \frac{1}{2}\sum_{k=1}^{c} \sum_{i,j\in I_k, i\neq j} \|o_i - o_j\|. \end{aligned} \tag{22}$$

Similarly, the sum of the between cluster distances is defined by:

$$\begin{aligned} S_B &= \sum_{k<\ell}^{c} \sum_{i\in I_k, j\in I_\ell} \|o_i - o_j\| \\ &= \frac{1}{2}\sum_{k\neq\ell}^{c} \sum_{i\in I_k, j\in I_\ell} \|o_i - o_j\|. \end{aligned} \tag{23}$$

Hence, the total sum of the distances is defined by:

$$S_T = \sum_{i<j} \|o_i - o_j\| = S_W + S_B. \tag{24}$$

In what follows, we recall the 27 crisp clustering indices mentioned in Ref. 2.

2.2.1. *Ball-Hall Index BHI*

The Ball-Hall index is the mean of clusterwize mean dispersions through all the clusters, given by

$$BHI = \frac{1}{c}\sum_{k=1}^{c} \frac{1}{n_k} \sum_{i\in I_k} \left|\left| o_i - z^{\{k\}} \right|\right|^2. \tag{25}$$

2.2.2. *Banfeld-Raftery Index BRI*

This index is the weighted sum of the logarithms of the within group scatter dispersion $WGSS^{\{k\}}$ divided by n_k:

$$BRI = \sum_{k=1}^{c} n_k \log \frac{WGSS^{\{k\}}}{n_k}. \tag{26}$$

2.2.3. *C Index CI*

The C Index is defined as:

$$CI = \frac{S_W - S_{min}}{S_{max} - S_{min}}. \tag{27}$$

where S_{min} is the sum of the N_W smallest distances of all pairs of distinct data and S_{max} is the sum of the N_W largest distances of all pairs of distinct data.[13]

2.2.4. *Calinski-Harabasz Index CHI*

The Calinski-Harabasz index is defined by the ratio of the unbiased between group scatter dispersion $BGSS/(c-1)$ to the unbiased sum of the within group scatter dispersions $WGSS/(n-c)$:

$$CHI = \frac{n-c}{c-1}\frac{BGSS}{WGSS}. \tag{28}$$

2.2.5. *Davies-Bouldin Index DBI*

The Davies Bouldin Index is the mean value, among all the clusters C_k, of the maximum M_k of the quotient $\frac{\delta_k + \delta_\ell}{\Delta_{k\ell}}$ for all $\ell \neq k$. Namely, DBI is defined by:

$$DBI = \frac{1}{c}\sum_{k=1}^{c} \max_{\ell\neq k} \left(\frac{\delta_k + \delta_\ell}{\Delta_{k\ell}}\right), \tag{29}$$

where

$$\delta_k = \frac{1}{n_k}\sum_{i\in I_k} \|o_i^{\{k\}} - z^{\{k\}}\| \tag{30}$$

and

$$\Delta_{k\ell} = d(z^{\{k\}}, z^{\{\ell\}}) = \|z^{\{\ell\}} - z^{\{k\}}\|. \tag{31}$$

2.2.6. *Det Ratio Index DRI*

The Det Ratio Index is defined as the quotient of the determinant of the scatter matrix T by the determinant of the sum of the within group scatter matrices WG:

$$DRI = \frac{\det(T)}{\det(WG)}. \tag{32}$$

2.2.7. *Dunn Index DI*

The Dunn Index is the ratio of the smallest distance between distinct clusters to the largest diameter of cluster, given as

$$DI = \frac{d_{min}}{d_{max}}, \tag{33}$$

where

$$d_{min} = \min_{k\neq\ell}\left(\min_{\substack{i\in I_k\\ j\in I_\ell}} \|o_i^{\{k\}} - o_j^{\{\ell\}}\|\right), \tag{34}$$

$$d_{max} = \max_{1\leq k\leq c}\left(\max_{\substack{i,j\in I_k\\ i\neq j}} \|o_i^{\{k\}} - o_j^{\{k\}}\|\right). \tag{35}$$

2.2.8. *Baker-Hubert Gamma Index BHGI*

The Baker-Hubert Gamma Index is an adaptation of the Γ-index[11] of correlation between two vectors of the same size.

Suppose we have two vectors $A = \{a_1, a_2, \ldots, a_n\}, B = \{b_1, b_2, \ldots, b_n\}$. Then, for two given distinct indices i, j, the two vectors are said to be concordant at the pair of indices $\{i, j\}$ if whenever $a_i < a_j$ and $b_i < b_j$ or $a_i > a_j$ and $b_i > b_j$. Otherwise, the two vectors are said to be discordant. When we denoted the number of concordant pairs by s^+, and the number of discordant pairs by s^-, the Γ-index is defined by

$$\Gamma = \frac{s^+ - s^-}{s^+ + s^-}. \tag{36}$$

In the context of clustering, we define the first vector to be the set of distances between two distinct points in the data (regardless of whether they are in the same cluster or not). The corresponding element of the second vector is binary. It is 0 if the two points are in the same cluster, and 1 otherwise. Then the Baker Hubert Gamma Index is defined exactly like Eq. (36):

$$BHGI = \frac{s^+ - s^-}{s^+ + s^-}. \tag{37}$$

In this case, s^+ is the number of times that a distance between two points in the same cluster is less than a distance between two points in different clusters. For a concordant pair, in this case, $b_i < b_j$ means that the i-th pair in the vector A is in the same cluster (i.e., $b_i = 0$) and the j-th pair in the vector A had two points in different clusters (i.e., $b_j = 1$ and thus $b_i = 0 < 1 = b_j$). The number s^- is the number of times the opposite situation occurs.

2.2.9. *The GD Index GDI*

The GD Index is the abbreviation of Generalized Dunn's Index. This index is used for evaluating the between clusters and the within clusters distances. The *GDI* Index is defined like this:

$$GDI = \frac{\min_{k\neq\ell}\delta(C_k, C_\ell)}{\max_k \Delta(C_k)}, \tag{38}$$

where δ measures the between-cluster distance and Δ measures the within-cluster distance with $1 \leq k \leq c$ and $1 \leq \ell \leq c$.

There are three different definitions of Δ and six definitions of δ. The definitions of the within cluster distances Δ are

$$\Delta_1(C_k) = \max_{\substack{i,j\in I_k\\ i\neq j}} d(o_i, o_j) \tag{39}$$

$$\Delta_2(C_k) = \frac{1}{n_k(n_k-1)} \sum_{\substack{i,j\in I_k\\ i\neq j}} d(o_i, o_j) \tag{40}$$

$$\Delta_3(C_k) = \frac{2}{n_k} \sum_{i\in I_k} d(o_i, z^{\{k\}}), \tag{41}$$

where, d is the Euclidean distance. And the definitions of the between cluster distances δ are:

$$\delta_1(C_k, C_\ell) = \min_{\substack{i\in I_k\\ j\in I_\ell}} d(o_i, o_j) \tag{42}$$

$$\delta_2(C_k, C_\ell) = \max_{\substack{i\in I_k\\ j\in I_\ell}} d(o_i, o_j) \tag{43}$$

$$\delta_3(C_k, C_\ell) = \frac{1}{n_k n_\ell} \sum_{\substack{i\in I_k\\ j\in I_\ell}} d(o_i, o_j) \tag{44}$$

$$\delta_4(C_k, C_\ell) = d(z^{\{k\}}, z^{\{\ell\}}) \tag{45}$$

$$\delta_5(C_k, C_\ell) = \frac{1}{n_k + n_\ell} \left(\sum_{i\in I_k} d(o_i, z^{\{\ell\}}) + \sum_{j\in I_\ell} d(o_j, z^{\{\ell\}})\right) \tag{46}$$

$$\delta_6(C_k, C_\ell) = \max\left\{\sup_{i\in I_k}\inf_{j\in I_\ell} d(o_i, o_j), \sup_{j\in I_\ell}\inf_{i\in I_k} d(o_i, o_j)\right\}. \tag{47}$$

The distances δ_1 to δ_4 are called single linkage clustering, complete linkage clustering, average linkage clustering, centroid linkage clustering respectively. The distance δ_5 is the weighted mean of the mean distances between the points in clusters C_k and C_ℓ with respect to their cluster center. The distance δ_6 is the Hausdorff distance.

2.2.10. *G Plus Index GPI*

By using the same notations in *Baker Hubert Gamma Index BHGI* in subsection 2.2.8, the $G+$ index is given as

$$GPI = \frac{s^-}{N_T(N_T-1)/2} = \frac{2s^-}{N_T(N_T-1)}. \quad (48)$$

2.2.11. *Ksq DetW Index KDWI*

The Ksq DetW Index is defined as the square of the number of clusters c times the determinant of the sum of the within group scatter matrices WG:

$$KDWI = c^2 \det(WG). \quad (49)$$

2.2.12. *Log Det Ratio Index LDRI*

The Log Det Ratio Index is the number of data n times the logarithm of the Det Ratio Index DRI (32):

$$LDRI = n \log DRI = n \log \frac{\det(T)}{\det(WG)}. \quad (50)$$

2.2.13. *Log SS Ratio Index LSSRI*

The Log SS Ratio Index is given by the logarithm of the ratio of the between group scatter dispersion $BGSS$ to the sum of the within group scatter dispersions $WGSS$:

$$LSSRI = \log \frac{BGSS}{WGSS}. \quad (51)$$

2.2.14. *The McClain-Rao Index McRI*

The McClaino-Rao Index is defined as the ratio of the mean within cluster distances S_W/N_W to the mean between cluster distances S_B/N_B:

$$McRI = \frac{S_W/N_W}{S_B/N_B} = \frac{N_B}{N_W}\frac{S_W}{S_B}. \quad (52)$$

2.2.15. *PBM Index PBMI*

The PBM Index uses the distances between the data and their cluster centers and also uses the distances between the cluster centers themselves. The $PBMI$ is defined by

$$PBMI = \left(\frac{1}{c} \times \frac{E_T}{E_W} \times D_B\right)^2, \quad (53)$$

where D_B is the maximum distance between two cluster centers, E_W is the sum of the distances of the data in each cluster to the center and E_T is the sum of the distances of all the data to the center in the whole data set, respectively given as follows:

$$D_B = \max_{k<\ell}\left(z^{\{k\}}, z^{\{\ell\}}\right), \quad (54)$$

$$E_W = \sum_{k=1}^{c}\sum_{i\in I_k} d(o_i, z^{\{k\}}), \quad (55)$$

$$E_T = \sum_{i=1}^{n} d(o_i, z). \quad (56)$$

2.2.16. *Point-Biserial Index PBI*

The Point-Biserial Index is essentially defined as the point-biserial coefficient of the distances of pairs of points and the homogeneity of pairs. For precise meaning of "homogeneity", see Ref. 2. PBI is defined by

$$PBI = \left(\frac{S_W}{N_W} - \frac{S_B}{N_B}\right)\frac{\sqrt{N_W N_B}}{N_T}. \quad (57)$$

2.2.17. *Ratkowsky-Lance Index RLI*

The Ratkowsky-Lance Index is defined as the square root of the mean R of attribute-wize ratios, say the p-th attribute ($1 \le p \le d$), of the between group scatter dispersion $BGSS_p$ to the total scattering TSS_p divided by the cluster number. Namely RLI is given by

$$RLI = \sqrt{\frac{R}{c}} = \sqrt{\frac{1}{cd}\sum_{p=1}^{d}\frac{BGSS_p}{TSS_p}}, \quad (58)$$

where

$$R = \frac{1}{d}\sum_{p=1}^{d}\frac{BGSS_p}{TSS_p}. \quad (59)$$

TSS_p and $BGSS_p$ are obtained by replacing the norms of vectors by the absolute value of difference of p attributes in TSS (12) and $BGSS$ (18), respectively.

2.2.18. *Ray-Turi Index RTI*

The Ray-Turi Index is given like this

$$RTI = \frac{1}{N}\frac{WGSS}{\min_{k<\ell}\Delta_{k\ell}^2}, \quad (60)$$

where $\Delta_{k\ell}$ is the minimum of the square distances between all cluster centers $\min_{k<\ell}\Delta_{k\ell}^2 = \min_{k<\ell}\|z^{\{k\}} - z^{\{\ell\}}\|^2$.

2.2.19. *Scott-Symons Index SSI*

The Scott-Symons Index is the weighted sum of the logarithms of the determinants of the variance-covariance matrix of each cluster:

$$SSI = \sum_{k=1}^{c} n_k \log \det\Big(\frac{WG^{\{k\}}}{n_k}\Big). \tag{61}$$

2.2.20. *SD Index SDI*

The SD Index consists of $\mathcal{S}$ and $\mathcal{D}$. $\mathcal{S}$ is the average scattering for clusters and $\mathcal{D}$ is the total separation between clusters. The quantity S is given by

$$\mathcal{S} = \frac{\frac{1}{c}\sum_{k=1}^{c}\|V^{\{k\}}\|}{\|V\|}, \tag{62}$$

where $V = {}^t(\mathrm{Var}(V_1), \ldots, \mathrm{Var}(V_d))$ is the variance vector with size d. Likewise $V^{\{k\}} = {}^t\Big(\mathrm{Var}\Big(V_1^{\{k\}}\Big), \ldots, \mathrm{Var}\Big(V_d^{\{k\}}\Big)\Big)$ is the variance vector with size d for each cluster C_k. The variance (Var) in this index refers to the usual statistical measurement. Moreover, the quantity $\mathcal{D}$ is given by

$$\mathcal{D} = \frac{\max\limits_{k<\ell}\|z^{\{k\}} - z^{\{\ell\}}\|}{\min\limits_{k<\ell}\|z^{\{k\}} - z^{\{\ell\}}\|}\sum_{k=1}^{c}\frac{1}{\sum_{\substack{\ell=1\\ \ell\neq k}}^{c}\|z^{\{k\}} - z^{\{\ell\}}\|}. \tag{63}$$

Hence, the SDI is defined like this

$$SDI = \alpha\mathcal{S} + \mathcal{D}, \tag{64}$$

where α is the weight assigned to the clustering that has the greatest number of clusters based on $\mathcal{D}$.

2.2.21. *S_Dbw Index SDbwI*

The S_Dbw Index consists of two items, within cluster variance $\mathcal{S}$ and between cluster density $\mathcal{G}$, given as

$$SDbwI = \mathcal{S} + \mathcal{G}, \tag{65}$$

where $\mathcal{S}$ is found in Eq. (62) and $\mathcal{G}$ is given as

$$\mathcal{G} = \frac{2}{c(c-1)}\sum_{k<\ell} R_{k\ell}, \tag{66}$$

where $R_{k\ell}$ is the quotient between the density at the midpoint and the largest density at the two centers:

$$R_{k\ell} = \frac{\gamma_{k\ell}(H_{k\ell})}{\max\big(\gamma_{k\ell}(z^{\{k\}}), \gamma_{k\ell}(z^{\{\ell\}})\big)}, \tag{67}$$

where, $\gamma_{k\ell}$ is the density for a given data point in the cluster C_k, C_ℓ. It is equal to the number of data point in these two cluster which the distance to this data point is less than the limit value σ:

$$\sigma = \frac{1}{c}\sqrt{\sum_{k=1}^{c}\|V^{\{k\}}\|}. \tag{68}$$

And $H_{k\ell}$ is the midpoint of the centers $z^{\{k\}}$ and $z^{\{\ell\}}$.

2.2.22. *Silhouette Index SI*

The Silhouette Index is calculated using the mean within cluster distance a and the smallest of the mean distances of the other clusters from each data b. The step in calculating the Silhouette Index starts by finding the mean distance of the i-th data o_i in cluster C_k with all the other data o_j in C_k which denoted by $a(i)$:

$$a(i) = \frac{1}{n_k - 1}\sum_{\substack{j\in I_k\\ j\neq i}} d(o_i, o_j), \tag{69}$$

and next we calculate the mean distance of $o_i \in C_k$ to each data in the different cluster C_ℓ ($\ell \neq k$):

$$\mathfrak{d}(o_i, C_\ell) = \frac{1}{n_\ell}\sum_{j\in I_\ell} d(o_i, o_j). \tag{70}$$

After we calculate $\mathfrak{d}(o_i, C_\ell)$ for all clusters C_ℓ ($\ell \neq k$), next we choose the smallest of the mean distances denoted by $b(i)$:

$$b(i) = \min_{\ell\neq k}\mathfrak{d}(o_i, C_\ell). \tag{71}$$

Then the Silhouette width is calculated by

$$s(i) = \frac{b(i) - a(i)}{\max\big(a(i), b(i)\big)}. \tag{72}$$

The value of $s(i)$ is between -1 and 1, with value near to 1 indicates the data o_i is affected to the appropriate clusters. On the other hand, the value near -1 indicates the data o_i should be affected to another cluster.

Let us denote $\mathfrak{s}_k$ as the mean of all silhoutte widths of data in the cluster C_k:

$$\mathfrak{s}_k = \frac{1}{n_k}\sum_{i\in I_k} s(i). \tag{73}$$

Then we obtain the global Silhoutte Index as the mean of the mean silhoutte for the clusters:

$$SI = \frac{1}{c}\sum_{k=1}^{c}\mathfrak{s}_k. \tag{74}$$

2.2.23. *Tau Index TI*

Using the same notations in the section 2.2.8, the Tau Index is defined like this

$$TI = \frac{s^+ - s^-}{\sqrt{N_B N_W \left(\frac{N_T(N_T-1)}{2}\right)}}. \tag{75}$$

2.2.24. *Trace W Index TrWI*

The Trace W Index is none other than the sum of the within group scatter dispersions:

$$TrWI = WGSS. \tag{76}$$

2.2.25. *Trace WiB Index TrWiBI*

The Trace WiB Index is defined by the trace of the inverse matrix of the sum of the within group scatter matrices WG times the between group scatter matrix BG:

$$TrWiBI = \mathrm{Tr}(WG^{-1}BG). \tag{77}$$

2.2.26. *Wemmert-Gançarsky Index WGI*

The Wemmert-Gançarsky Index is calculated by using the distances between the data and the centers of all the clusters.

Let us denote $R(o)$ as the quotient of distance of the data o to the center of the cluster C_k to which the data belongs and the smallest distance of the data to the center of all the other clusters:

$$R(o) = \frac{\|o - z^{\{k\}}\|}{\min\limits_{\ell \neq k} \|o - z^{\{\ell\}}\|}. \tag{78}$$

The mean of these quotients is then calculated for each cluster. If the result is greater than 1, then it is ignored. Otherwise its complement to 1 is calculated. It is done by using the quantity below:

$$J_k = \max\left\{0, 1 - \frac{1}{n_k}\sum_{i \in I_k} R(o_i)\right\}. \tag{79}$$

Hence, The Wemmert-Gançarsky Index is expressed as the weighted mean of J_k for all the clusters:

$$WGI = \frac{1}{n}\sum_{k=1}^{c} n_k J_k. \tag{80}$$

This expression can be rewritten as follows

$$WGI = \frac{1}{n}\sum_{k=1}^{c} \max\left\{0, n_k - \sum_{i \in I_k} R(o_i)\right\}, \tag{81}$$

2.2.27. *Xie-Beni Index XBI*

This index is also known as an index for fuzzy clustering. But it can be used for crisp clustering. This index is given as

$$XBI = \frac{1}{n}\frac{WGSS}{\min\limits_{k \neq \ell} \delta_1(C_k, C_\ell)}, \tag{82}$$

where δ measures the between cluster distances (cf. δ_1 Eq. (43) in section 2.2.9).

3. Fuzzification

Our fuzzification is essentially done by introducing membership degrees, but in adapting this to various indices, we disaggregate it into its components and build up the fuzzification. It is done in the following order. First, as a preliminary step, statistics using powers of membership are introduced (3.1). This makes it possible to deal with the distribution of the data in each cluster or in the whole. Apart from this, we can also deal with the distribution of each data over clusters. Under these statistical treatments, the various statistics needed for fuzzifying the index are expressed in terms of membership degrees (3.2). Finally, under these preparations, the fuzzification of individual indices is performed (3.3).

3.1. *Statistical treatment of μ^m*

Fuzzification is a process of converting the crisp quantities with values in $\{0, 1\}$ to fuzzy ones with values in $[0, 1]$ through the exponentiated membership degree μ_{ik}^m. In our approach, instead of μ_{ik} itself, we consider the exponentiated quantity μ_{ik}^m as the ratio of occurrence of the data o_i in the cluster C_k. Hence the total probability of the occurrence of the data o_i is given as:[1]

$$p_i = \frac{\sum_{k=1}^{c} \mu_{ik}^m}{\sum_{j=1}^{n}\sum_{\ell=1}^{c} \mu_{j\ell}^m}. \tag{83}$$

Likewise in the cluster C_k, the within cluster probability of the i-th data o_i is given as:

$$p_i^{\{k\}} = \frac{\mu_{ik}^m}{\sum_{j=1}^{n} \mu_{jk}^m}. \tag{84}$$

On the other hand, the classification probability of the i-th data o_i which occurs into the k-th cluster C_k, given as:

$$p_{ik} = \frac{\mu_{ik}^m}{\sum_{\ell=1}^{c} \mu_{i\ell}^m}. \tag{85}$$

The difference between the within cluster distribution and the classification distribution is illustrated in Fig.1.

Let $f(o)$ be a function of data o. Then its expectations with respect to the total expectation and the within cluster expectation in the cluster C_k, respectively given as:

$$\mathrm{E}[f(o)] = \sum_{i=1}^{n} p_i f(o_i) \tag{86}$$

and

$$\mathrm{E}^{\{k\}}[f(o)] = \sum_{i=1}^{n} p_i^{\{k\}} f(o_i). \tag{87}$$

o_1	μ_{11}^m	$\cdots$	μ_{1k}^m	$\cdots$ μ_{1c}^m
$\vdots$	$\vdots$		$\vdots$	$\vdots$
o_i	μ_{i1}^m	$\cdots$	μ_{ik}^m	$\cdots$ μ_{ic}^m
$\vdots$	$\vdots$		$\vdots$	$\vdots$
o_n	μ_{n1}^m	$\cdots$	μ_{nk}^m	$\cdots$ μ_{nc}^m
	C_1	$\cdots$	C_k	$\cdots$ C_c

(a) the within cluster probability

o_1	μ_{11}^m	$\cdots$	μ_{1k}^m	$\cdots$	μ_{1c}^m
$\vdots$	$\vdots$		$\vdots$		$\vdots$
o_i	μ_{i1}^m	$\cdots$	μ_{ik}^m	$\cdots$	μ_{ic}^m
$\vdots$	$\vdots$		$\vdots$		$\vdots$
o_n	μ_{n1}^m	$\cdots$	μ_{nk}^m	$\cdots$	μ_{nc}^m
	C_1	$\cdots$	C_k	$\cdots$	C_c

(b) the classification probability

Fig. 1: Within cluster probability vs. classification probability

For the sake of convenience, we introduce the following quantities that are the total weighted sum of the quantity f of data o, given as:

$$\varpi[f(o)] = \sum_{k=1}^{c}\sum_{i=1}^{n} \mu_{ik}^m f(o_i) \tag{88}$$

and the within cluster weighted sum in cluster C_k of the quantity f of the data o, given as:

$$\varpi^{\{k\}}[f(o)] = \sum_{i=1}^{n} \mu_{ik}^m f(o_i). \tag{89}$$

Hence, we have the total population as:

$$\nu = \varpi[1] \tag{90}$$

and the within cluster population of the cluster C_k as:

$$\nu^{\{k\}} = \varpi^{\{k\}}[1]. \tag{91}$$

In the fuzzification proposed in this article, we use the notions of total/within-cluster populations instead of total/within-cluster numbers. Immediately from Eq. (1), (2), (3) and the condition $m > 1$, we have

$$0 < \nu^{\{k\}} < \nu \leq n \qquad (1 \leq k \leq c). \tag{92}$$

It is easily verified that in (92), the equality $\nu = n$ holds iff the clustering is crisp, i.e., $\mu_{ik} \in \{0, 1\}$. Hence, the expectations are rewritten as follows:

$$\mathrm{E}[f(o)] = \frac{\varpi[f(o)]}{\nu} \tag{93}$$

and

$$\mathrm{E}^{\{k\}}[f(o)] = \frac{\varpi^{\{k\}}[f(o)]}{\nu^{\{k\}}}. \tag{94}$$

3.2. *Fuzzification of Basic Quantities*

3.2.1. *Barycenter and Parametric Center*

In Fuzzy C-Means or K-Means where representative points for data sets are used, formal representative points or the center of gravity of the data are often employed. Furthermore, when the center of gravity is employed, several metrics are used to measure the distance between clusters. One of them is Euclidean distance. In this article, the center of gravity of the cluster is calculated as the *barycenter*. And apart from this, we also use the words *parametric centers* for representative points of clusters.

The barycenter of all data is denoted by b. And as the center of gravity of whole data, the barycenter is equal to the expectation of data, namely:[1]

$$b = \mathrm{E}[o]. \tag{95}$$

Likewise the barycenter of data within the k-th cluster C_k is denoted by $b^{\{k\}}$ which is equal to the expectation of data in the k-th cluster C_k:

$$b^{\{k\}} = \mathrm{E}^{\{k\}}[o]. \tag{96}$$

Furthermore, the total parametric center is denoted by z which is equal to the barycenter of all data, namely:

$$z = b, \tag{97}$$

as well as the within cluster parametric center in the k-th cluster C_k is denoted by $z^{\{k\}}$ which is equal to the barycenter of the k-th cluster C_k:

$$z^{\{k\}} = b^{\{k\}}. \tag{98}$$

3.2.2. *Variance-Covariance Matrix*

Similarly to the scatter matrix T and the within group scattering matrix $WG^{\{k\}}$ in crisp clustering, in fuzzy clustering, the variance-covariance matrix measures the dispersion of the membership degree which involves the expectation of the data to their centers. Here we use Σ as the covariance matrix of total data and $\Sigma^{\{k\}}$ as the one of data in the k-th cluster C_k, respectively given as:

$$\Sigma = \mathrm{E}\Big[(o-z)\ {}^t(o-z)\Big] \tag{99}$$

and

$$\Sigma^{\{k\}} = \mathrm{E}^{\{k\}}\Big[(o-z^{\{k\}})\ {}^t(o-z^{\{k\}})\Big]. \tag{100}$$

Finally, let us introduce a third matrix $\Delta^{\{k\}}$ called the k-th cluster displacement dyad that has important relationships with these two matrices. This matrix measures the displacement of the center $z^{\{k\}}$ of the k-th cluster with respect to the total center z given by:

$$\Delta^{\{k\}} = (z^{\{k\}}-z)\ {}^t(z^{\{k\}}-z) \tag{101}$$

It can be easily verified that the matrices $\Sigma, \Sigma^{\{k\}}$ and $\Delta^{\{k\}}$ are related as

$$\begin{aligned} &\mathrm{E}\big[(o-z^{\{k\}})\ {}^t(o-z^{\{k\}})\big] = \Sigma + \Delta^{\{k\}}, \\ &\mathrm{E}^{\{k\}}\big[(o-z)\ {}^t(o-z)\big] = \Sigma^{\{k\}} + \Delta^{\{k\}}. \end{aligned} \tag{102}$$

3.3. *Fuzzification of Crisp Clustering Indices*

In this subsection, firstly we fuzzify the key quantities given in 2.2 (a) - (g) and secondly we fuzzify the 27 crisp clustering indices:

(a) The Scatter Matrix T (11) can be fuzzified as:

$$\begin{aligned} T &= \nu\Sigma \\ &= \varpi[(o_i-z)\ {}^t(o_i-z)]. \end{aligned} \tag{103}$$

where ν and Σ are found in (90) and (99) respectively.

(b) The Total Scattering TSS (12) can be fuzzified as:

$$\begin{aligned} TSS &= \nu\mathrm{Tr}\Sigma \\ &= \varpi[\|o-z\|^2]. \end{aligned} \tag{104}$$

(c) The Within Group Scatter $WG^{\{k\}}$ (13) can be fuzzified as:

$$\begin{aligned} WG^{\{k\}} &= \nu^{\{k\}}\Sigma^{\{k\}} \\ &= \varpi^{\{k\}}[(o-z^{\{k\}})\ {}^t(o-z^{\{k\}})]. \end{aligned} \tag{105}$$

where $\nu^{\{k\}}\Sigma^{\{k\}}$ is found in (91) and (100) respectively.

And the fuzzification of their sum of all the clusters WG (14) is:

$$\begin{aligned} WG &= \sum_{k=1}^{c} \nu^{\{k\}}\Sigma^{\{k\}} \\ &= \sum_{k=1}^{c} \varpi^{\{k\}}[(o-z^{\{k\}})\ {}^t(o-z^{\{k\}})]. \end{aligned} \tag{106}$$

(d) The Within Scatter Dispersion $WGSS^{\{k\}}$ (15), can be fuzzified as:

$$\begin{aligned} WGSS^{\{k\}} &= \nu^{\{k\}}\mathrm{Tr}\Sigma^{\{k\}} \\ &= \varpi^{\{k\}}[||o-z^{\{k\}}||^2]. \end{aligned} \tag{107}$$

And the fuzzification of the sum of the within group scatter dispersion $WGSS$ (16) is given as:

$$\begin{aligned} WGSS &= \sum_{k=1}^{c} \nu^{\{k\}}\mathrm{Tr}\Sigma^{\{k\}} \\ &= \sum_{k=1}^{c} \varpi^{\{k\}}[||o-z^{\{k\}})||^2]. \end{aligned} \tag{108}$$

(e) The Between Group Scatter BG (17) is given in fuzzy form as:

$$BG = \sum_{k=1}^{c} \nu^{\{k\}}\triangle^{\{k\}}, \tag{109}$$

where $\triangle^{\{k\}}$ is found in (101).

(f) The Between Group Scatter Dispersion $BGSS$ (18) can be fuzzified as:

$$BGSS = \sum_{k=1}^{c} \nu^{\{k\}}\|z^{\{k\}}-z\|^2. \tag{110}$$

(g) Pairs of Point. To fuzzify these pairwise quantities, recall that the i-th data o_i has the exponentiated membership degree μ_{ik}^m in the cluster C_k. Thus the distribution of μ_{ik}^m over clusters leads to the classification probability (85). Hence, the probability that the i-th data o_i occurs in the k-th cluster C_k and the the j-th data o_j occurs in the t-th cluster C_t at the same time is as follows:

$$p_{ik}p_{jt} = \frac{\mu_{ik}^m}{\sum_{\ell=1}^{c}\mu_{i\ell}^m}\frac{\mu_{jt}^m}{\sum_{s=1}^{c}\mu_{js}^m}. \tag{111}$$

Now define the within cluster means of $f(o,o')$ which occurs for data o_i and o_j:

$$\omega_{\{i,j\}}\big[f(o,o')\big] = \sum_{k=1}^{c} p_{ik}p_{jk}f(o_i,o_j) \tag{112}$$

and the between cluster means of $f(o,o')$ which occurs for data o_i and o_j:

$$\beta_{\{i,j\}}\big[f(o,o')\big] = \sum_{k\neq\ell} p_{ik}p_{j\ell}f(o_i,o_j), \quad (113)$$

then the total population for co-clusterity of data o_i and o_j is

$$\omega_{\{i,j\}}[1] = \sum_{k=1}^{c} p_{ik}p_{jk}, \quad (114)$$

and the total population for anti-clusterity of data o_i and o_j is

$$\beta_{\{i,j\}}[1] = \sum_{k=1}^{c} p_{ik}p_{jk}. \quad (115)$$

In both case, co-clusterity and anty-clusterity are mutually complementary:

$$\text{co-clusterity} + \text{anti-clusterity} = 1.$$

Finally we define

$$\omega\big[f(o,o')\big] = \frac{1}{2}\sum_{i\neq j}\omega_{\{i,j\}}f(o_i,o_j) \quad (116)$$

and

$$\beta\big[f(o,o')\big] = \frac{1}{2}\sum_{i\neq j}\beta_{\{i,j\}}f(o_i,o_j). \quad (117)$$

Then the total population of pairs of distinct points within cluster N_W (19) and the population of pairs of points which do not belong to the same cluster N_B (20) respectively can be fuzzified as:

$$N_W = \omega[1], \quad (118)$$

$$N_B = \beta[1]. \quad (119)$$

On the other hand, the fuzzification of the sum of the within cluster distance S_W (22)) and the sum of the between cluster distances S_B (23) respectively given as:

$$S_W = \omega[\|o-o'\|]. \quad (120)$$

$$S_B = \beta[\|o-o'\|]. \quad (121)$$

We note that, in fuzzy forms, the dependency of $\omega_{\{i,j\}}\big[f(o,o')\big]+\beta_{\{i,j\}}\big[f(o,o')\big]$ and $\omega\big[f(o,o')\big]+\beta\big[f(o,o')\big]$ on μ's are purely caused by that of the function f. In fact we have that

This follows from the equality:

$$\omega_{\{i,j\}}\big[f(o,o')\big] + \beta_{\{i,j\}}\big[f(o,o')\big] = f(o_i,o_j), \quad (122)$$

and

$$\omega\big[f(o,o')\big] + \beta\big[f(o,o')\big] = \frac{1}{2}\sum_{i\neq j} f(o_i,o_j). \quad (123)$$

In particular, $N_T = N_W + N_B$ is equal to the total population of pairs of distinct data and $S_T = S_W + S_B$ is equal to the sum of all distances between pair of distinct data. Note that both of N_T and S_T are independent of μ's.

Now we are in the position to fuzzify the 27 crisp clustering indices.

3.3.1. *Ball-Hall Index BHI*

The Ball-Hall Index (25) can be fuzzified as:

$$BHI = \frac{1}{c}\sum_{k=1}^{c}\mathrm{E}^{\{k\}}\big[\|o-z^{\{k\}}\|^2\big]. \quad (124)$$

3.3.2. *Banfeld-Raftery Index BRI*

The Banfeld-Raftery Index (26) can be fuzzified as:

$$BRI = \sum_{k=1}^{c}\nu^{\{k\}}\,\log\left(\mathrm{E}^{\{k\}}\big[\|o-z^{\{k\}}\|^2\big]\right). \quad (125)$$

This fuzzification is derived from (107).

3.3.3. *C Index CI*

To fuzzify this index, we introduce the order statistics to find the minimum or maximum value of a sample in the probability cases. We denote S as a set of statistic samples and S_a is the a-th smallest value of order statistic of S ($1 \le a \le |S|$); $S^a = -((-S)_a)$ is the a-th largest value of S. Hence, we get

$$S_{min} = \sum_{a=1}^{N_W} S_a, \quad (126)$$

$$S_{max} = \sum_{a=1}^{N_W} S^a. \quad (127)$$

Then C Index (27) can be fuzzified as

$$CI = \frac{\omega[\|o-o'\|] - S_{min}}{S_{max} - S_{min}} \quad (128)$$

This fuzzification is derived from (120).

3.3.4. *Calinski-Harabasz Index CHI*

The Calinski-Harabasz Index (28) can be fuzzified as:

$$CHI = \frac{n-c}{c-1} \left(\frac{\varpi\big[\|o-z\|^2\big]}{\sum_{k=1}^{c} \varpi^{\{k\}}\big[\|o-z^{\{k\}}\|^2\big]} - 1 \right). \tag{129}$$

This fuzzification is derived from (110) and (108).

3.3.5. *Davis-Bouldin Index DBI*

The Davis-Bouldin Index (29) can be fuzzified as

$$DBI = \frac{1}{c} \sum_{k=1}^{c} \max_{k \neq \ell} \left(\frac{\mathrm{E}^{\{k\}}[\|o-z^{\{k\}}\|] + \mathrm{E}^{\{\ell\}}[\|o-z^{\{\ell\}}\|]}{\|z^{\{k\}} - z^{\{\ell\}}\|} \right) \tag{130}$$

3.3.6. *Det Ratio Index DRI*

The Det Ratio Index (32) can be fuzzified like this

$$DRI = \frac{\det \varpi\big[(o-z)\ {}^t(o-z)\big]}{\det\left(\sum_{k=1}^{c} \varpi^{\{k\}}\big[(o-z^{\{k\}})\ {}^t(o-z^{\{k\}})\big] \right)}. \tag{131}$$

This fuzzification is derived from (103) and (106).

3.3.7. *Dunn Index DI*

The Dunn Index (33) can be fuzzified like this

$$DI = \frac{\min_{k\neq\ell}\left(\min_{i\neq j}\left(p_{ik}p_{j\ell}[\|o_i - o_j\|] \right) \right)}{\max_{k}\left(\max_{i\neq j}\left(p_{ik}p_{jk}[\|o_i - o_j\|] \right) \right)} \tag{132}$$

3.3.8. *Baker-Hubert Gamma Index BHGI*

There are two main components in this index such as the concordant pairs and the discordant pairs. For any pair of pairs $\{i,j\}(i \neq j)$ and $(i',j')(i' \neq j')$, for a concordant pairs iff $\big((\|o_i - o_j\| < \|o_{j'} - o_{i'}\|$ and $(\beta_{\{i,j\}}[1] < \beta_{\{i',j'\}}[1])\big)$ or $\big((\|o_i - o_j\| > \|o_{i'} - o_{j'}\|)$ and $(\beta_{\{i,j\}}[1] > \beta_{\{i',j'\}}[1])\big)$. For discordant pairs, iff $\big((\|o_i - o_j\| < \|o_{j'} - o_{i'}\|$ and $(\beta_{\{i,j\}}[1] > \beta_{\{i',j'\}}[1])\big)$ or $\big((\|o_i - o_j\| > \|o_{i'} - o_{j'}\|)$ and $(\beta_{\{i,j\}}[1] < \beta_{\{i',j'\}}[1])\big)$. Then the number of concordant pairs are denoted by s^+ and the number of discordant pairs are denoted by s^-. Hence the fuzzification of this index is

$$BHGI = \frac{s^+ - s^-}{s^+ + s^-} \tag{133}$$

where

$$s^+ = \#\{\{\{i,j\},\{i',j'\}\}|\{i,j\} \text{ and } \{'i,j'\} \text{ are concordant}\} \tag{134}$$

and

$$s^- = \#\{\{\{i,j\},\{i',j'\}\}|\{i,j\} \text{ and } \{'i,j'\} \text{ are discordant}\} \tag{135}$$

which # refers to the cardinality of set.

3.3.9. *GDI Index*

There are 3 definitions of the within cluster distances Δ that can be fuzzified as follows

$$\Delta_1(C_k) = \max_{i\neq j} p_{ik}p_{jk}[\|o_i - o_j\|] \tag{136}$$

$$\Delta_2(C_k) = \frac{1}{\nu^{\{k\}}\big(\nu^{\{k\}}-1\big)} p_{ik}p_{jk}\|o_i - o_j\| \tag{137}$$

$$\Delta_3(C_k) = 2\mathrm{E}^{\{k\}}[\|o - z^{\{k\}}\|]. \tag{138}$$

And there are six definitions of the between cluster distances δ that can be fuzzified as follows

$$\delta_1(C_k, C_\ell) = \min_{i\neq j}\left(p_{ik}p_{j\ell}[\|o_i - o_j\|] \right) \tag{139}$$

$$\delta_2(C_k, C_\ell) = \max_{i\neq j}\left(p_{ik}p_{j\ell}[\|o_i - o_j\|] \right) \tag{140}$$

$$\delta_3(C_k, C_\ell) = \sum_{i\neq j} p_{ik}p_{j\ell}[\|o_i - o_j\|] \tag{141}$$

$$\delta_4(C_k, C_\ell) = [\|z^{\{k\}} - z^{\{\ell\}}\|] \tag{142}$$

$$\delta_5(C_k, C_\ell) = \frac{1}{\nu^{\{k\}} + \nu^{\{\ell\}}} \left(\sum_{k=1}^{c} \varpi^{\{k\}}[\|o - z^{\{k\}}\|] + \sum_{\ell=1}^{c} \varpi^{\{\ell\}}[\|o - z^{\{\ell\}}\|] \right) \tag{143}$$

$$\delta_6(C_k, C_\ell) = \max\{\sup_i \inf_j p_{ik}p_{j\ell}||o_i - o_j\|, \sup_j \inf_i p_{ik}p_{j\ell}||o_i - o_j\|\} \tag{144}$$

3.3.10. *G Plus Index GPI*

The G-plus Index (48) can be fuzzified as

$$GPI = \frac{2s^-}{n(n-1)\Big(\big(n(n-1)\big)-1\Big)}, \tag{145}$$

where s^- is found in (135).

3.3.11. *The Ksq DetW Index KDWI*

The Ksq DetW Index (49) can be fuzzified as:

$$KDWI = c^2 \det \Big(\sum_{k=1}^{c} \varpi^{\{k\}}\big[(o-z^{\{k\}})\ {}^t(o-z^{\{k\}})\big]\Big). \tag{146}$$

This fuzzification is derived from (106).

3.3.12. *Log Det Ratio Index LDRI*

The Log Det Ratio Index(50) can be fuzzified as:

$$LDRI = \nu \log \frac{\det\ \varpi\big[(o-z)\ {}^t(o-z)\big]}{\det\ \Big(\sum_{k=1}^{c} \varpi^{\{k\}}\big[(o-z^{\{k\}})\ {}^t(o-z^{\{k\}})\big]\Big)}. \tag{147}$$

This fuzzification is derived from (103) and (106).

3.3.13. *Log SS Ratio Index LSSRI*

The Log SS Ratio Index (51) can be fuzzified as:

$$LSSRI = \log\Big(\frac{\varpi\big[\|o-z\|^2\big]}{\sum_{k=1}^{c}\varpi^{\{k\}}\big[\|o-z^{\{k\}}\|^2\big]}-1\Big). \tag{148}$$

This fuzzification is derived from (108) and (110).

3.3.14. *McClain-Rao Index McRI*

The McClain-Rao Index (52) can be fuzzified like this:

$$McRI = \frac{\sum_{i<j}\sum_{k\neq\ell} p_{ik}p_{j\ell}}{\sum_{i<j}\sum_{k=1}^{c} p_{ik}p_{jk}} \frac{\sum_{i<j}\sum_{k=1}^{c} p_{ik}p_{jk}\|o_i-o_j\|}{\sum_{i<j}\sum_{k\neq\ell} p_{ik}p_{j\ell}\|o_i-o_j\|}. \tag{149}$$

This form can also be written as:

$$McRI = \frac{\beta[1]\omega[\|o-o'\|]}{\omega[1]\beta[\|o-o'\|]}, \tag{150}$$

where β and ω are found in (112) and (113), respectively. And this fuzzification is derived from (118), (119), (120) and (121).

3.3.15. *PBM Index PBMI*

In this index, there are several components that must be fuzzified. Firstly, D_B (54) can be fuzzified as

$$D_B = \max_{k<\ell}\|z^{\{k\}}-z^{\{\ell\}}\|. \tag{151}$$

Secondly, E_W (55) can be fuzzified as

$$E_W = \varpi[\|o-z^{\{k\}}\|]. \tag{152}$$

Thirdly, E_T (56) can be fuzzified as

$$E_T = \varpi[\|o-z\|]. \tag{153}$$

Hence $PBMI$ (53) is defined as

$$PBMI = \Big(\frac{1}{c}\times\frac{E_T}{E_W}\times D_B\Big)^2 \tag{154}$$

or can be rewritten as

$$PBMI = \Big(\frac{1}{c}\times\frac{\varpi[\|o-z\|]}{\varpi[\|o-z^{\{k\}}\|]}\times \big(\max_{k<\ell}\|z^{\{k\}}-z^{\{\ell\}}\|\big)\Big)^2 \tag{155}$$

3.3.16. *Point-Biserial Index PBI*

The Point-Biserial Index (57) can be fuzzified as:

$$PBI = \frac{2}{n(n-1)} \Big(\omega[\|o-o'\|]\sqrt{\frac{\beta[1]}{\omega[1]}}-\beta[\|o-o'\|]\sqrt{\frac{\omega[1]}{\beta[1]}}\Big). \tag{156}$$

This fuzzification is derived from (118), (119), (120) and (121).

3.3.17. *Ratkowsky-Lance Index RLI*

The Ratkowsky-Lance Index (58) can be fuzzified as:

$$RLI^2 = \frac{1}{cd}\sum_{p=1}^{d}\frac{\nu^{\{k\}}\Delta_{pp}^{\{k\}}}{\nu\Sigma_{pp}}, \tag{157}$$

or this index can be written also as:

$$RLI^2 = \frac{1}{cd}\sum_{p=1}^{d}\frac{\sum_{k=1}^{c}\nu^{\{k\}}(z^{\{k\}p}-z^p)^2}{\nu\mathrm{E}\big[(o_*^p-z^p)^2\big]}. \tag{158}$$

Here o_*^p indicates the p-th row (attribute) component of data variable o.

3.3.18. *Ray-Turi Index RTI*

The Ray-Turi Index RTI (60) can be fuzzified as

$$RTI = \frac{1}{n}\frac{\varpi^{\{k\}}[|||o - z^{\{k\}}||^2]}{\min_{k<\ell}\Delta^2_{k\ell}} \tag{159}$$

This fuzzification is derived from (107) and $\min_{k<\ell}\Delta^2_{k\ell} = \min_{k<\ell}\|z^{\{k\}} - z^{\{\ell\}}\|$.

3.3.19. *Scott-Symons Index SSI*

The Scott-Symons Index (61) can be fuzzified as:

$$SSI = \sum_{k=1}^{c}\nu \log\det\Big(\frac{\varpi^{\{k\}}\big[(o - z^{\{k\}})\ {}^t(o - z^{\{k\}})\big]}{\nu^{\{k\}}}\Big). \tag{160}$$

This fuzzification is derived from (105).

3.3.20. *SD Index SDI*

There are two quantities that must be fuzzified in this index such as $\mathcal{S}$ and $\mathcal{D}$. The quantity $\mathcal{S}$ in (62) can be fuzzified as

$$\mathcal{S} = \frac{\frac{1}{c}\sum_{k=1}^{c}\sqrt{\sum_{p=1}^{d}\big(\mathrm{E}^{\{k\}}[(o_p - z_p^{\{k\}})^2]\big)^2}}{\sqrt{\sum_{p=1}^{d}\big(\mathrm{E}[(o_p - z_p)^2]\big)^2}}, \tag{161}$$

and the quantity $\mathcal{D}$ in (63) can be fuzzified as

$$\mathcal{D} = \frac{\max\limits_{k\neq\ell}\|z^{\{k\}} - z^{\{\ell\}}\|}{\min\limits_{k\neq\ell}\|z^{\{k\}} - z^{\{\ell\}}\|}\sum_{k=1}^{c}\frac{1}{\sum_{\ell\neq k}\|z^{\{k\}} - z^{\{\ell\}}\|} \tag{162}$$

which these quantities is derived from (142). Hence the SD Index (64) can be fuzzified as:

$$SDI = \alpha\mathcal{S} + \mathcal{D} \tag{163}$$

where α is the weight assigned to the clustering that has the greatest number of clusters based on $\mathcal{D}$.

3.3.21. *S_Dbw Index SDbwI*

To fuzzify this index, first, we fuzzify the limit value σ (68), given as

$$\sigma = \frac{1}{c}\sum_{k=1}^{c}\sqrt{\sum_{p=1}^{d}\Big(\mathrm{E}^{\{k\}}[(o_p - z_p^{\{k\}})^2]\Big)^2}, \tag{164}$$

then the density $\gamma_{k\ell}(o)$ given as

$$\gamma_{k\ell}(o) = \Big(\sum_{i \text{ with } \|o_i - o\|<\sigma} p_{ik}\Big) + \Big(\sum_{i \text{ with } \|o_i - o\|<\sigma} p_{i\ell}\Big) \tag{165}$$

and then the quotient $R_{k\ell}$ (67) can be fuzzified as

$$R_{k\ell} = \frac{\gamma_{k\ell}\Big(\frac{z^{\{k\}} + z^{\{\ell\}}}{2}\Big)}{\max\{\gamma_{k\ell}(z^{\{k\}}), \gamma_{k\ell}(z^{\{\ell\}})\}}. \tag{166}$$

Finally, the between cluster density $\mathcal{G}$ (66) can be fuzzified as

$$\mathcal{G} = \frac{1}{c(c-1)}\sum_{k\neq\ell}R_{k\ell}. \tag{167}$$

By using the Eqs.(161), (167), the *S_Dbw Index* in fuzzy form given as

$$SDbwI = \mathcal{S} + \mathcal{G}. \tag{168}$$

3.3.22. *Silhouette Index SI*

In this index, first of all, we fuzzify the within cluster mean distance $a(i)$ (69) as

$$a(i) = \min\{\mathrm{E}[\|o - o_i\|] \quad | \quad 1 \le k \le c\}, \tag{169}$$

then, the mean distance $\mathfrak{d}(o_i, C_\ell)$ (70), given as

$$\mathfrak{d}(o_i, C_\ell) = \mathrm{E}^{\{\ell\}}[\|o - o_i\|], \tag{170}$$

and finally the smallest of the mean distances $b(i)$ (71) as

$$b(i) = \{\mathrm{E}^{\{k\}}[\|o - o_i\|] \quad | \quad 1 \le k \le c\}_2, \tag{171}$$

where the suffix "2" indicates that this quantity is the second order statistic. Then by using (169) and (171), the fuzzification of the Silhouette width $s(i)$ (72) of o_i is given by

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}. \tag{172}$$

Hence, the mean of silhouette width $\mathfrak{s}_k$ (73) can be fuzzified as

$$\mathfrak{s}_k = \mathrm{E}^{\{k\}}[s(o)], \tag{173}$$

where $s(o)$ is the function of objects defined $s(o_i) = s(i)$. And finally, by using (173) the global sillhouette index (74) can be fuzzified as

$$SI = \frac{1}{c}\sum_{k=1}^{c}\mathfrak{s}_k. \tag{174}$$

3.3.23. *Tau Index TI*

The Tau Index $TrWI$ can be fuzzified as

$$TI = s^+ - s^- \Big/ \sqrt{\beta[1]\omega[1]\Big(\frac{n(n-1)\big((n(n-1))-1\big)}{2}\Big)}, \tag{175}$$

where this fuzzification is derived from (118), (119), (134) and (135).

3.3.24. *Trace W Index TrWI*

The Trace W Index (76) can be fuzzified as

$$TWI = \sum_{k=1}^{c} \varpi^{\{k\}}\big[\big\|(o - z^{\{k\}})\big\|^2\big]. \tag{176}$$

This fuzzification is derived from (107).

3.3.25. *Trace WiB Index TrWiBI*

The Trace WiB Index (77) can be fuzzified as

$$TrWiBI = \mathrm{Tr}\Big(\big(\sum_{k=1}^{c}\nu^{\{k\}}\Sigma^{\{k\}}\big)^{-1}\big(\sum_{k=1}^{c}\nu^{\{k\}}\Delta^{\{k\}}\big)\Big), \tag{177}$$

or we can write as:

$$TrWiBI = \sum_{k=1}^{c}\nu^{\{k\}}\|z^{\{k\}} - z\|^2_{WG^{-1}}, \tag{178}$$

where $\|\ \|_{WG^{-1}}$ is norm on R^d induced from the symmetric matrix WG^{-1}. The fuzzification is derived from (106) and (109).

3.3.26. *Wemmert-Gançarsky Index WGI*

In this index, the quotient $R(o)$ (78) can be fuzzified as

$$R(o) = \frac{\min\{\|o - z^{\{k\}}\|] \mid 1 \le k \le c\}}{\{\|o - z^{\{k\}}\|] \mid 1 \le k \le c\}_2}, \tag{179}$$

then J_k (79) can be fuzzified as

$$J_k = \max\{0, 1 - \mathrm{E}^{\{k\}}R(o)\}. \tag{180}$$

Hence by using (180), *The Wemmert-Gançarsky Index* can be fuzzified as

$$WGI = \frac{1}{n}\sum_{k=1}^{c}\nu^{\{k\}}J_k. \tag{181}$$

3.3.27. *Xie-Beni Index XBI*

The Xie-Beni Index (82) can be fuzzified as

$$XBI = \frac{\sum_{k=1}^{c}\varpi^{\{k\}}[\|o - z^{\{k\}}\|^2]}{\min\limits_{k<\ell}\Big(\min\limits_{i\neq j}\big(p_{ik}p_{j\ell}[\|o_i - o_j\|]\big)\Big)}. \tag{182}$$

The fuzzification is derrived from (108) and (139).

4. Optimization of Clustering Indices

As mentioned in the previous section that in Ref. 2 there are 13 analytic indices and 14 non-analytic indices. In this article, we restrict our attention to the analytic ones to be optimized by using the gradient method, while this method can not be used to the non-analytic indices. General optimization problems involving non-analytic functions are widely treated in mathematical programming, and many methods have been proposed.[14,16,17] However for this problem, we keep it as our next project. Hence in this section, we only discuss about the optimization on 13 analytic indices by using the gradient method.

Gradient ∇ is a differential operator that gives a vector field with all partial derivatives as components to a multivariate function, giving the direction of maximum increase for that function at each point. It is used as the search direction at the current point of consideration to solve the problem of maximization or minimization of a function.

In FCM, the objective function J (10) is minimized with the search directions for membership degree μ_{ik} and for parametric centers z and $z^{\{k\}}$. The steps of FCM algorithm are as follows[15] :

(1) Initialize the membership degree randomly $\mu_{ik}(0)$
(2) Calculate the centers $z^{\{k\}}$
(3) Calculate the membership degree $\mu_{ik}(1)$
(4) Optimize the Objective Function J under the constraints:

$$\sum_{k=1}^{c}\mu_{ik} = 1. \tag{183}$$

(5) Repeat the step 2-4 until the stopping conditions are achieved, namely the maximum iteration and the error value.

In FCM, the objective function is limited by constraint thus we use also the technique of La-

grange multiplier for solving constrained optimization problems. The Lagrange multipliers come from the gradient vectors of the objective function and the constrained function at a point is perpendicular to the graph of the function which passed through that point.[20] The common form of Lagrangian function is:[18]

$$L(x, \lambda_r) = f(x) - \sum_r \lambda_r g_r(x),$$

where $f(x)$ is an objective function, x is the original multivariables, λ_r are Lagrange multipliers. And the Lagrangian function in FCM is given as

$$L(\mu, z^{\{k\}}, \lambda) = J - \sum_{i=1}^{n} \lambda_i g_i,$$

for which the optimality conditions are:

$$\nabla_\mu L(\mu, z^{\{k\}}, \lambda) = 0,$$
$$\nabla_{z^{\{k\}}} L(\mu, z^{\{k\}}, \lambda) = 0,$$
$$\nabla_{\lambda_i} L(\mu, z^{\{k\}}, \lambda) = g_i = 0.$$

When the objective function includes the total centers z, the following equation is added to the above.

$$\nabla_z L(\mu, z, \lambda) = 0.$$

4.1. *Optimization of Ball-Hall Index (BHI)*

4.1.1. *Parametric Representative*

Regarding the parametric within cluster center $z^{\{k\}}$'s as the representative point of clusters, the FCM problem is to minimize the objective function of BHI given as

$$BHI = \frac{1}{c} \sum_{k=1}^{c} \mathrm{E}^{\{k\}} \big[\|o - z^{\{k\}}\|^2 \big].$$

By using the Lagrange multiplier, we can write the Lagrangian function as

$$L(\mu, z^{\{k\}}, \lambda) = BHI - \sum_{i=1}^{n} \lambda_i g_i, \tag{184}$$

where $g_i = \sum_{\ell=1}^{c} \mu_{i\ell} - 1$. Taking the first order partial derivatives with respect to $\mu, z^{\{k\}}$ and λ and putting them equal to zero, respectively we get

$$\frac{\partial}{\partial \mu_{ik}} \left(BHI - \sum_{i=1}^{n} \lambda_i g_i \right) = 0 \qquad (1 \le i \le n, 1 \le k \le c), \tag{185}$$

$$\frac{\partial}{\partial z^{\{k\}p}} \left(BHI - \sum_{i=1}^{n} \lambda_i g_i \right) = 0 \qquad (1 \le k \le c, 1 \le p \le d), \tag{186}$$

$$\frac{\partial}{\partial \lambda} \left(BHI - \sum_{i=1}^{n} \lambda_i g_i \right) = 0 \qquad (1 \le i \le n). \tag{187}$$

Eqs. (185), (186) and (187) respectively imply that

$$\frac{1}{c} \frac{m \mu_{ik}^{m-1}}{\nu^{\{k\}}} \left(\|z^{\{k\}} - o_i\|^2 - \mathrm{E}^{\{k\}} \left[\|z^{\{k\}} - o\|^2 \right] \right) = \lambda_i, \tag{188}$$

$$\frac{2}{c} \, {}^t e_p \big(z^{\{k\}} - b^{\{k\}} \big) = 0, \tag{189}$$

$$g_i = 0. \tag{190}$$

Where e_p is the p-th member of the standard basis of R^d:

$${}^t e_p = \begin{matrix} p \\ \vee \\ 0, ..., 0, 1, 0, ..., 0. \end{matrix}$$

Eqs. (190) provide the original constraints and Eqs. (189) are equivalent to

$${}^t e_p z^{\{k\}} = {}^t e_p b^{\{k\}}.$$

Since ${}^t e_p$'s span the dual space of the d-dimensional data space, we have

$$z^{\{k\}} = b^{\{k\}}.$$

At the extremal point (of variables), the parametric within cluster center $z^{\{k\}}$ must be the barycenter $b^{\{k\}}$, thus we have

$$z^{\{k\}} = b^{\{k\}} = \mathrm{E}^{\{k\}}[o]. \tag{191}$$

And the (188) can be rewritten as

$$\frac{1}{c} \frac{m \mu_{ik}^{m-1}}{\nu^{\{k\}}} \left(\|b^{\{k\}} - o_i\|^2 - V^{\{k\}} \right) = \lambda_i, \tag{192}$$

where

$$V^{\{k\}} = \mathrm{E}^{\{k\}} \big[\|b^{\{k\}} - o\|^2 \big].$$

Hence we get μ_{ik} as

$$\mu_{ik} = \left(\frac{c\lambda_i}{m} \right)^{\frac{1}{m-1}} \left(\frac{\|b^{\{k\}} - o_i\|^2 - V^{\{k\}}}{\nu^{\{k\}}} \right)^{-\frac{1}{m-1}}. \tag{193}$$

Then by substituting the (193) into (190), we have

$$1 = \left(\frac{c\lambda_i}{m}\right)^{\frac{1}{m-1}} \sum_{\ell=1}^{c} \left(\frac{\|b^{\{\ell\}} - o_i\|^2 - V^{\{\ell\}}}{\nu^{\{\ell\}}}\right)^{-\frac{1}{m-1}}. \quad (194)$$

Hence we get the membership degree as

$$\mu_{ik} = \frac{\left(\frac{\|b^{\{k\}} - o_i\|^2 - V^{\{k\}}}{\nu^{\{k\}}}\right)^{-\frac{1}{m-1}}}{\sum_{\ell=1}^{c}\left(\frac{\|b^{\{\ell\}} - o_i\|^2 - V^{\{\ell\}}}{\nu^{\{\ell\}}}\right)^{-\frac{1}{m-1}}} \quad (195)$$

The cluster center updating equation is calculated in (191) and the membership degree updating equation is calculated in (195).

4.1.2. *Barycentric Representatives*

By applying $z^{\{k\}}$'s as the barycenters of clusters, BHI can be rewritten as

$$BHI = BHI(\mu) = \frac{1}{c}\sum_{k=1}^{c} \mathrm{E}^{\{k\}}\big[\|o - b^{\{k\}}\|^2\big].$$

Then the Lagrange function can be written as

$$L(\mu, \lambda) = BHI - \sum_{i=1}^{n} \lambda_i g_i, \quad (196)$$

where $g_i = \sum_{\ell=1}^{c} \mu_{i\ell} - 1$. Taking the first order partial derivatives with respect to μ, λ and putting them equal to zero, respectively we get the similar results as (185) and (187). Now $z^{\{k\}}$'s are omitted because they are no longer independent variables. Furthermore, by direct calculation, the equations (185) and (187) are respectively expressed as

$$\frac{1}{c}\frac{m\mu_{ik}^{m-1}}{\nu^{\{k\}}}\left(\|b^{\{k\}} - o_i\|^2 - V^{\{k\}}\right) = \lambda_i, \quad (197)$$

$$g_i = 0. \quad (198)$$

The systems of equations under the conditions (191) are exactly same as (188) and (190). Hence, our discussion of these problems are same as the case for the parametric centers.

4.2. *Optimization of Banfeld-Raftery Index (BRI)*

Regarding the parametric within cluster center $z^{\{k\}}$'s as the representative point of clusters, the FCM problem is to minimize the objective function of BRI, given as:

$$BRI = \sum_{k=1}^{c} \nu^{\{k\}} \log\big(\mathrm{E}^{\{k\}}\big[\|o - z^{\{k\}}\|^2\big]\big).$$

By using the Lagrange multiplier, we can write the Lagrangian function as

$$L(\mu, z^{\{k\}}, \lambda) = BRI - \sum_{i=1}^{n} \lambda_i g_i, \quad (199)$$

where $g_i = \sum_{\ell=1}^{c} \mu_{i\ell} - 1$. Then taking the first order partial derivatives with respect to $\mu, z^{\{k\}}$ and λ and putting them equal to zero, respectively we get

$$\frac{\partial}{\partial \mu_{ik}}\left(BRI - \sum_{j=1}^{n}\lambda_j g_j\right) = 0 \quad (1 \le i \le n, 1 \le k \le c), \quad (200)$$

$$\frac{\partial}{\partial z^{\{k\}p}}\left(BRI - \sum_{j=1}^{n}\lambda_j g_j\right) = 0 \quad (1 \le k \le c, 1 \le p \le d), \quad (201)$$

$$\frac{\partial}{\partial \lambda}\left(BRI - \sum_{j=1}^{n}\lambda_j g_j\right) = 0 \qquad (1 \le i \le n). \quad (202)$$

Eqs. (200), (201) and (202) respectively imply that

$$m\mu_{ik}^{m-1}\left(\log\big(\mathrm{E}^{\{k\}}\big[\|z^{\{k\}} - o\|^2\big]\big) + \frac{\|z^{\{k\}} - o_i\|^2}{\mathrm{E}^{\{k\}}\big[\|z^{\{k\}} - o\|^2\big]}\right) = \lambda_i, \quad (203)$$

$$2\nu^{\{k\}}\frac{{}^t e_p\big(z^{\{k\}} - b^{\{k\}}\big)}{\mathrm{E}^{\{k\}}\big[\|z^{\{k\}} - o\|^2\big]} = 0, \quad (204)$$

$$g_i = 0. \quad (205)$$

Eqs. (205) provide the original constraints and Eqs. (204) are equivalent to

$$\nu^{\{k\}}\frac{{}^t e_p z^{\{k\}}}{\mathrm{E}^{\{k\}}\big[\|z^{\{k\}} - o\|^2\big]} = \nu^{\{k\}}\frac{{}^t e_p b^{\{k\}}}{\mathrm{E}^{\{k\}}\big[\|z^{\{k\}} - o\|^2\big]}. \quad (206)$$

And hence, these equations cause the parametric within cluster center must be the barycenter

$$z^{\{k\}} = b^{\{k\}} = \mathrm{E}^{\{k\}}[o]. \quad (207)$$

And (203) are rewritten as

$$m\mu_{ik}^{m-1}\left(\log V^{\{k\}} + \frac{\|b^{\{k\}} - o_i\|^2}{V^{\{k\}}} - 1\right) = \lambda_i. \quad (208)$$

Eq. (208) can be solved with respect to μ' as

$$\mu_{ik} = \left(\frac{\lambda_i}{m}\right)^{\frac{1}{m-1}} \left(\log V^{\{k\}} + \frac{\|b^{\{k\}} - o_i\|^2}{V^{\{k\}}} - 1\right)^{-\frac{1}{m-1}}, \tag{209}$$

then substituting this into (183), we have:

$$1 = \left(\frac{\lambda_i}{m}\right)^{\frac{1}{m-1}} \sum_{\ell=1}^{c} \left(\log V^{\{k\}} + \frac{\|b^{\{k\}} - o_i\|^2}{V^{\{k\}}} - 1\right)^{-\frac{1}{m-1}}. \tag{210}$$

Hence, we get the membership degree as follows:

$$\mu_{ik} = \left(\log V^{\{k\}} + \frac{\|b^{\{k\}} - o_i\|^2}{V^{\{k\}}} - 1\right)^{-\frac{1}{m-1}} \Bigg/ \sum_{\ell=1}^{c} \left(\log V^{\{\ell\}} + \frac{\|b^{\{\ell\}} - o_i\|^2}{V^{\{\ell\}}} - 1\right)^{-\frac{1}{m-1}}. \tag{211}$$

The cluster center updating equation is calculated in Eq. (207) and the membership degree updating equations is calculated in Eq. (211) .

The Barycentric representative in this index is similar discussion to the Ball Hall Index (BHI).

4.3. *Optimization of Calinski-Harabasz Index (CHI)*

Regarding parametric total center z and the parametric within cluster centers $z^{\{k\}}$ as the representative points of all data and data in clusters, respectively, the FCM problem is to minimize the objective function of CHI, given by

$$CHI = \frac{n-c}{c-1}\left(\frac{\varpi\left[\|o-z\|^2\right]}{\sum_{k=1}^{c}\varpi^{\{k\}}\left[\|o-z^{\{k\}}\|^2\right]} - 1\right).$$

In what follows we consider the function

$$\frac{\varpi\left[||z-o||^2\right]}{\sum_{k=1}^{c}\varpi^{\{k\}}\left[||z^{\{k\}}-o||^2\right]} \tag{212}$$

as an objective function CHI instead of original CHI since these optimization problems are equivalent. By using the Lagrange multiplier, we can write the Lagrangian function as

$$L(\mu, z, z^{\{k\}}, \lambda) = CHI - \sum_{i=1}^{n}\lambda_i g_i \tag{213}$$

where $g_i = \sum_{\ell=1}^{c}\mu_{i\ell} - 1$. Then taking the first order partial derivatives with respect to $\mu, z, z^{\{k\}}$ and λ and putting them equal to zero, respectively we get

$$\frac{\partial}{\partial \mu_{ik}}\left(CHI - \sum_{j=1}^{n}\lambda_j g_j\right) = 0 \quad (1 \le i \le n, 1 \le k \le c), \tag{214}$$

$$\frac{\partial}{\partial z^p}\left(CHI - \sum_{j=1}^{n}\lambda_j g_j\right) = 0 \quad (1 \le p \le d), \tag{215}$$

$$\frac{\partial}{\partial z^{\{k\}p}}\left(CHI - \sum_{j=1}^{n}\lambda_j g_j\right) = 0 \quad (1 \le k \le c, 1 \le p \le d), \tag{216}$$

$$\frac{\partial}{\partial \lambda}\left(CHI - \sum_{j=1}^{n}\lambda_j g_j\right) = 0 \qquad (1 \le i \le n). \tag{217}$$

Eqs. (214),(215), (216) and (217), respectively imply that

$$m\mu_{ik}^{m-1}\frac{\|z-o_i\|^2 - CHI\|z^{\{k\}}-o_i\|^2}{\sum_{\ell=1}^{c}\varpi^{\{\ell\}}\left[\|z^{\{\ell\}}-o\|^2\right]} = \lambda_i, \tag{218}$$

$$\frac{2\nu\ {}^t e_p(z-b)}{\sum_{\ell=1}^{c}\varpi^{\{\ell\}}\left[\|z^{\{\ell\}}-o\|^2\right]} = 0, \tag{219}$$

$$-CHI\frac{2\nu^{\{k\}}\ {}^t e_p(z^{\{k\}}-b^{\{k\}})}{\sum_{\ell=1}^{c}\varpi^{\{\ell\}}\left[\|z^{\{\ell\}}-o\|^2\right]} = 0, \tag{220}$$

$$g_i = 0. \tag{221}$$

Eqs. (221) provide the original constraints and as usual (219) and (220) cause

$$z = b = \mathrm{E}[o], \tag{222}$$

$$z^{\{k\}} = b^{\{k\}} = \mathrm{E}^{\{k\}}[o], \tag{223}$$

provided that $CHI \neq 0$. Then (218) can be rewritten as

$$m\mu_{ik}^{m-1}\frac{\|b-o_i\|^2 - CHI\|b^{\{k\}}-o_i\|^2}{\sum_{\ell=1}^{c}\nu^{\{\ell\}}V^{\{\ell\}}} = \lambda_i \tag{224}$$

and then can be solved with respect to μ's as

$$\mu_{ik} = \left(\frac{\lambda_i\sum_{\ell=1}^{c}\nu^{\{\ell\}}V^{\{\ell\}}}{m\|b-o_i\|^2}\right)^{\frac{1}{m-1}} \left(1 - CHI\frac{\|b^{\{k\}}-o_i\|^2}{\|b-o_i\|^2}\right)^{-\frac{1}{m-1}} \tag{225}$$

then by substituting (225) into (183), we have

$$1=\left(\frac{\lambda_i\sum_{\ell=1}^{c}\nu^{\{\ell\}}V^{\{\ell\}}}{m\|b-o_i\|^2}\right)^{-\frac{1}{m-1}} \sum_{k=1}^{c}\left(1-CHI\frac{\|b^{\{k\}}-o_i\|^2}{\|b-o_i\|^2}\right)^{-\frac{1}{m-1}}. \tag{226}$$

Hence we get the equations for membership degree updating rule as

$$\mu_{ik}=\frac{\left(1-CHI\frac{\|b^{\{k\}}-o_i\|^2}{\|b-o_i\|^2}\right)^{-\frac{1}{m-1}}}{\sum_{\ell=1}^{c}\left(1-CHI\frac{\|b^{\{\ell\}}-o_i\|^2}{\|b-o_i\|^2}\right)^{-\frac{1}{m-1}}}. \tag{227}$$

The total cluster centers updating equations are calculated by (222) and (223), respectively and the membership degrees updating equations are calculated by Eq. (227).

The barycentric representative in this index is similar discussion to the previous index.

4.4. *Optimization of Det Ratio Index (DRI)*

Regarding the parametric total center z and the parametric within cluster centers $z^{\{k\}}$ as the representative points of all data and data in clusters, respectively, the FCM problem is to minimize the objective function of DRI, given as

$$DRI=\frac{\det\varpi\big[(o-z)\ {}^t(o-z)\big]}{\det\Big(\sum_{k=1}^{c}\varpi^{\{k\}}\big[(o-z^{\{k\}})\ {}^t(o-z^{\{k\}})\big]\Big)}. \tag{228}$$

By using the Lagrange multiplier, we can write the Lagrangian function as

$$L(\mu,z,\lambda)=DRI-\sum_{i=1}^{n}\lambda_i g_i, \tag{229}$$

where $g_i=\sum_{\ell=1}^{c}\mu_{i\ell}-1$. Then taking the first order partial derivatives with respect to $\mu, z, z^{\{k\}}$ and λ and putting them equal to zero, respectively we get

$$\frac{\partial}{\partial\mu_{ik}}\left(DRI-\sum_{i=1}^{n}\lambda_i g_i\right)=0 \quad (1\le i\le n, 1\le k\le c), \tag{230}$$

$$\frac{\partial}{\partial z^p}\left(DRI-\sum_{i=1}^{n}\lambda_i g_i\right)=0 \quad (1\le p\le d), \tag{231}$$

$$\frac{\partial}{z^{\{k\}p}}\left(DRI-\sum_{i=1}^{n}\lambda_i g_i\right)=0 \quad (1\le k\le c, 1\le p\le d), \tag{232}$$

$$\frac{\partial}{\partial\lambda}\left(DRI-\sum_{i=1}^{n}\lambda_i g_i\right)=0 \qquad (1\le i\le n). \tag{233}$$

Eqs. (230), (231), (232) and (233), respectively imply that

$$m\mu_{ik}^{m-1}DRI\big(\|o_i-z\|^2_{T^{-1}}-\|o_i-z^{\{k\}}\|^2_{WG^{-1}}\big)=\lambda_i, \tag{234}$$

$$-2\nu DRI\ {}^t e_p\ T^{-1}(b-z)=0, \tag{235}$$

$$-2\nu^{\{k\}}DRI\ {}^t e_p(WG)^{-1}(b^{\{k\}}-z^{\{k\}})=0, \tag{236}$$

$$g_i=0. \tag{237}$$

The norm $\|\ \|_{T^{-1}}$ in (234) is the norm on R^d induced from the symmetric matrix T^{-1}. Eq. (237) provide the original constraints, and (235), (236) cause

$$z=b=\mathrm{E}[o], \tag{238}$$

$$z^{\{k\}}=b^{\{k\}}=\mathrm{E}^{\{k\}}[o], \tag{239}$$

provided that $DRI\neq 0$. Hence similarly as usual, we have

$$z=b=\mathrm{E}[o], \tag{240}$$

$$z^{\{k\}}=b^{\{k\}}=\mathrm{E}^{\{k\}}[o], \tag{241}$$

And the Eq. (234) can be rewritten as

$$m\mu_{ik}^{m-1}DRI\big(\|o_i-b\|^2_{T^{-1}}-\|o_i-b^{\{k\}}\|^2_{WG^{-1}}\big)=\lambda_i \tag{242}$$

Eq. (242) are formally solved with respect to μ's as

$$\mu_{ik}=\left(\frac{c\lambda_i}{mDRI}\right)^{\frac{1}{m-1}}\left(\|o_i-b\|^2_{T^{-1}}-\|o_i-b^{\{k\}}\|^2_{WG^{-1}}\right)^{-\frac{1}{m-1}}. \tag{243}$$

Then substitute them into (183), we have

$$1=\left(\frac{c\lambda_i}{mDRI}\right)^{\frac{1}{m-1}}\sum_{\ell=1}^{c}\left(\|o_i-b\|^2_{T^{-1}}-\|o_i-b^{\{k\}}\|^2_{WG^{-1}}\right)^{-\frac{1}{m-1}}. \tag{244}$$

Hence, we have the membership degree updating rule as

$$\mu_{ik} = \left(\|o_i - b\|^2_{T^{-1}} - \|o_i - b^{\{k\}}\|^2_{WG^{-1}}\right)^{\frac{1}{m-1}} \Big/ \sum_{\ell=1}^{c}\left(\|o_i - b\|^2_{T^{-1}} - \|o_i - b^{\{\ell\}}\|^2_{WG^{-1}}\right)^{\frac{1}{m-1}}. \quad (245)$$

The total and the within cluster centers updating equations are calculated as (240), (241) and the membership degrees updating equations are calculated as (245).

The Barycentric representative in this index is similar discussion to the previous index.

4.5. *Optimization of Ksq DetW Index (KDWI)*

Regarding the parametric within cluster center $z^{\{k\}}$'s as the representative point of cluster, the FCM problem is to minimize the objective function of KDWI, given as

$$KDWI = c^2 \det\Big(\sum_{k=1}^{c} \varpi^{\{k\}}\big[(o - z^{\{k\}})\ {}^t(o - z^{\{k\}})\big]\Big). \quad (246)$$

By using the Lagrange multiplier, we can write the Lagrangian function as

$$L(\mu, z) = KDWI - \sum_{i=1}^{n} \lambda_i g_i, \quad (247)$$

where $g_i = \sum_{\ell=1}^{c} \mu_{i\ell} - 1$. Then taking the first order partial derivatives with respect to $\mu, z^{\{k\}}, \lambda$ are equal to zero, respectively we get:

$$\frac{\partial}{\partial \mu_{ik}}\left(KDWI - \sum_{i=1}^{n} \lambda_i g_i\right) = 0 \quad (1 \le i \le n, 1 \le k \le c), \quad (248)$$

$$\frac{\partial}{\partial z^{\{k\}p}}\left(KDWI - \sum_{i=1}^{n} \lambda_i g_i\right) = 0 \quad (1 \le k \le c, 1 \le p \le d), \quad (249)$$

$$\frac{\partial}{\partial \lambda_i}\left(KDWI - \sum_{i=1}^{n} \lambda_i g_i\right) = 0 \qquad (1 \le i \le n). \quad (250)$$

Eqs. (248), (249) and (250), respectively imply the followings:

$$m\mu_{ik}^{m-1}(KDWI)\|o_i - z^{\{k\}}\|^2_{WG^{-1}} = \lambda_i, \quad (251)$$

$$-2\nu^{\{k\}}(KDWI)^t e_p (WG)^{-1}(b^{\{k\}} - z^{\{k\}}) = 0, \quad (252)$$

$$g_i = 0. \quad (253)$$

Eqs. (253) provide the original constraints. And as usual, Eqs.(252) cause the parametric within cluster center must be the barycenter

$$z^{\{k\}} = b^{\{k\}} = \mathrm{E}^{\{k\}}[o]. \quad (254)$$

Hence (251) can be rewritten as:

$$m\mu_{ik}^{m-1}(KDWI)\|o_i - b^{\{k\}}\|^2_{WG^{-1}} = \lambda_i \quad (255)$$

and then μ_{ik} become

$$\mu_{ik} = \left(\frac{\lambda_i}{m(KDWI)}\right)^{\frac{1}{m-1}} \left(\|o_i - b^{\{k\}}\|^2_{WG^{-1}}\right)^{-\frac{1}{m-1}}. \quad (256)$$

And then substituting this into (183), we have

$$1 = \left(\frac{\lambda_i}{m(KDWI)}\right)^{\frac{1}{m-1}} \sum_{\ell=1}^{c}\left(\|o_i - b^{\{\ell\}}\|^2_{WG^{-1}}\right)^{-\frac{1}{m-1}}. \quad (257)$$

Hence, we get the membership degree as

$$\mu_{ik} = \frac{\left(\|o_i - z^{\{k\}}\|^2_{WG^{-1}}\right)^{-\frac{1}{m-1}}}{\sum_{\ell=1}^{c}\left(\|o_i - z^{\{\ell\}}\|^2_{WG^{-1}}\right)^{-\frac{1}{m-1}}} \quad (258)$$

The cluster center updating equation is calculated in (254) and the membership degree updating equation is calculated in (258).

The Barycentric representative in this index is similar discussion to the previous index

4.6. *Optimization of Log Det Ratio Index (LDRI)*

Since the logarithmic function log is monotonically increasing, the optimization problem for LDRI is equivalent to that for DRI, see section 4.4.

4.7. *Optimization of Log SS Ratio Index (LSSRI)*

Since the logarithmic function log is monotonically increasing, the optimization problem for LDRI is equivalent to that for CHI, see section 4.3.

4.8. *Optimization of McClain-Rao Index (McRI)*

The FCM problem to minimize the objective function of McRI, given as:

$$McRI = \frac{\beta[1]\omega[\|o-o'\|]}{\omega[1]\beta[\|o-o'\|]}.$$

By using the Lagrange multiplier, we can write the Langrangian function as

$$L(\mu,\lambda) = McRI - \sum_{i=1}^{n}\lambda_i g_i, \tag{259}$$

where $g_i = \sum_{\ell=1}^{c}\mu_{i\ell} - 1$. Taking the first order partial derivatives with respect to μ and λ and putting them equal to zero, respectively we have

$$\frac{\partial}{\partial\mu_{ik}}\left(McRI - \sum_{j=1}^{n}\lambda_j g_j\right) = 0, \tag{260}$$

$$\frac{\partial}{\partial\lambda_i}\left(McRI - \sum_{j=1}^{n}\lambda_j g_j\right) = 0. \tag{261}$$

Eqs. (260) and (261), respectively imply that

$$\begin{aligned}\frac{\mu_{ik}^{m-1}}{\nu_i}\sum_{\substack{1\le j\le n\\ j\neq i}}\left(p_{jk} - \sum_{\ell=1}^{c}p_{i\ell}p_{j\ell}\right)\\ \left(\frac{N_T}{N_B N_W} - \frac{S_T}{S_B S_W}\|o_i - o_j\|\right) = \lambda_i,\end{aligned} \tag{262}$$

$$g_i = 0, \tag{263}$$

where N_W, N_B, S_W and S_B respectively are found in (118), (119),(120) and (121).

Eq. (263) provide the original constraints and (262) are formally solved with respect to μ's as

$$\begin{aligned}\mu_{ik} = \\ \left(\frac{\lambda_i\nu_i}{m}\right)^{\frac{1}{m-1}}\left(\sum_{\substack{1\le j\le n\\ j\neq i}}\left(p_{jk} - \sum_{\ell=1}^{c}p_{i\ell}p_{j\ell}\right)\right.\\ \left.\times\left(\frac{N_T}{N_B N_W} - \frac{S_T}{S_B S_W}\|o_i - o_j\|\right)\right)^{-\frac{1}{m-1}}.\end{aligned} \tag{264}$$

Then substituting Eq. (264) into Eq. (183), we have

$$\begin{aligned}1 = \\ \left(\frac{\lambda_i\nu_i}{m}\right)^{\frac{1}{m-1}}\sum_{k=1}^{c}\left(\sum_{\substack{1\le j\le n\\ j\neq i}}\left(p_{jk} - \sum_{\ell=1}^{c}p_{i\ell}p_{j\ell}\right)\right.\\ \left.\times\left(\frac{N_T}{N_B N_W} - \frac{S_T}{S_B S_W}\|o_i - o_j\|\right)\right)^{-\frac{1}{m-1}}.\end{aligned} \tag{265}$$

Hence, we obtain the membership degree μ_{ik} as

$$\begin{aligned}\mu_{ik} = \left(\left(\sum_{\substack{1\le j\le n\\ j\neq i}}\left(p_{jk} - \sum_{\ell=1}^{c}p_{i\ell}p_{j\ell}\right)\right.\right.\\ \left.\times\left(\frac{N_T}{N_B N_W}\frac{S_T}{S_B S_W}\|o_i - o_j\|\right)\right)^{-\frac{1}{m-1}}\Bigg/\\ \left(\sum_{k=1}^{c}\left(\sum_{\substack{1\le j\le n\\ j\neq i}}\left(p_{jk} - \sum_{\ell=1}^{c}p_{i\ell}p_{j\ell}\right)\right)\right.\\ \left.\left.\times\left(\frac{N_T}{N_B N_W}\frac{S_T}{S_B S_W}\|o_i - o_j\|\right)\right)^{-\frac{1}{m-1}}\right)\end{aligned} \tag{266}$$

The membership degree updating rules are given by (266).

4.9. *Optimization of Point-Biserial Index (PBI)*

The FCM problem to minimize the objective function of PBI, given as

$$\begin{aligned}PBI = \frac{2}{n(n-1)}\left(\omega[\|o-o'\|]\sqrt{\frac{\beta[1]}{\omega[1]}}\right.\\ \left. - \beta[\|o-o'\|]\sqrt{\frac{\omega[1]}{\beta[1]}}\right).\end{aligned} \tag{267}$$

By using the Lagrange multiplier, we can write the Lagrange function as

$$L(\mu,\lambda) = PBI - \sum_{i=1}^{n}\sum_{i=1}^{n}\lambda_i g_i = 0, \tag{268}$$

where $g_i = \sum_{\ell=1}^{c}\mu_{i\ell} - 1$. Then taking the first order partial derivative with respect to μ and λ and putting them equal to zero, respectively given as

$$\frac{\partial}{\partial\mu_{ik}}\left(PBI - \sum_{j=1}^{n}\lambda_j g_j\right) = 0, \tag{269}$$

$$\frac{\partial}{\partial\lambda_i}\left(PBI - \sum_{j=1}^{n}\lambda_j g_j\right) = 0. \tag{270}$$

Eqs. (269) and Eq. (270), respectively imply that

$$\frac{m\mu_{ik}^{m-1}}{\nu_i}\frac{1}{\sqrt{N_W N_B}} \times \left(\frac{1}{2}\left(\frac{S_W}{N_W}+\frac{S_B}{N_B}\right)\sum_{\substack{1\le j\le n\\ j\neq i}}\left(p_{jk}-\sum_{\ell=1}^{c}p_{i\ell}p_{j\ell}\right) - \left(\sum_{\substack{1\le j\le n\\ j\neq i}}\left(p_{jk}-\sum_{\ell=1}^{c}p_{i\ell}p_{j\ell}\right)\|o_i-o_j\|\right)\right)=\lambda_i, \tag{271}$$

$$g_i = 0, \tag{272}$$

where N_W, N_B, S_W and S_B, respectively are found in (118), (119), (120) and (121).

Eq. (272) provide the original constraints and (271) are formally solved with respect to μ's as

$$\mu_{ik} = \left(\frac{\lambda_i\nu_i\sqrt{N_W N_B}}{m}\right)^{\frac{1}{m-1}} \times \left(\frac{1}{2}\left(\frac{S_W}{N_W}+\frac{S_B}{N_B}\right)\sum_{\substack{1\le j\le n\\ j\neq i}}\left(p_{jk}-\sum_{\ell=1}^{c}p_{i\ell}p_{j\ell}\right) - \left(\sum_{\substack{1\le j\le n\\ j\neq i}}\left(p_{jk}-\sum_{\ell=1}^{c}p_{i\ell}p_{j\ell}\right)\|o_i-o_j\|\right)\right)^{-\frac{1}{m-1}}. \tag{273}$$

Then substituting them into (183), we have

$$1 = \sum_{k=1}^{c}\left(\frac{\lambda_i\nu_i\sqrt{N_W N_B}}{m}\right)^{\frac{1}{m-1}} \times \left(\frac{1}{2}\left(\frac{S_W}{N_W}+\frac{S_B}{N_B}\right)\sum_{\substack{1\le j\le n\\ j\neq i}}\left(p_{jk}\sum_{\ell=1}^{c}p_{i\ell}p_{j\ell}\right) - \left(\sum_{\substack{1\le j\le n\\ j\neq i}}\left(p_{jk}-\sum_{\ell=1}^{c}p_{i\ell}p_{j\ell}\right)\|o_i-o_j\|\right)\right)^{-\frac{1}{m-1}}. \tag{274}$$

Hence we obtain the updating rule for membership degree μ_{ik} as

$$\mu_{ik} = \left(\frac{1}{2}\left(\frac{S_W}{N_W}+\frac{S_B}{N_B}\right)\sum_{\substack{1\le j\le n\\ j\neq i}}\left(p_{jk}-\sum_{\ell=1}^{c}p_{i\ell}p_{j\ell}\right) - \left(\sum_{\substack{1\le j\le n\\ j\neq i}}\left(p_{jk}-\sum_{\ell=1}^{c}p_{i\ell}p_{j\ell}\right)\|o_i-o_j\|\right)\right)^{-\frac{1}{m-1}} \Bigg/ \sum_{h=1}^{c}\left(\frac{1}{2}\left(\frac{S_W}{N_W}+\frac{S_B}{N_B}\right)\left(\sum_{\substack{1\le j\le n\\ j\neq i}}\left(p_{jh}-\sum_{\ell=1}^{c}p_{i\ell}p_{j\ell}\right)\right) - \left(\sum_{\substack{1\le j\le n\\ j\neq i}}\left(p_{jh}-\sum_{\ell=1}^{c}p_{i\ell}p_{j\ell}\right)\right)\|o_i=o_j\|\right)^{-\frac{1}{m-1}}. \tag{275}$$

The membership degree updating rules are given by (275)

4.10. *Optimization of Ratkowsky-Lance Index (RLI)*

4.10.1. *Parametric Representative*

Regarding parametric total center z and the parametric within cluster centers $z^{\{k\}}$ as the representative points of all data and data in clusters, respectively, the FCM problem is to minimize the objective function of RLI square instead of RLI itself, given as

$$RLI^2 = \frac{1}{cd}\sum_{p=1}^{d}\frac{\sum_{k=1}^{c}\nu^{\{k\}}(z^{\{k\}p}-z^p)^2}{\nu\mathrm{E}\left[(o_*^p-z^p)^2\right]}, \tag{276}$$

or we can write RLI^2 as

$$RLI^2 = \frac{1}{cd}\sum_{q=1}^{d}\frac{BGSS_q}{TSS_q}. \tag{277}$$

For a general differentiation $'$, we have

$$\begin{aligned}(RLI^2) &= \frac{1}{cd}\sum_{q=1}^{d}\left(\frac{BGSS_q}{TSS_q}\right)\\ &= \frac{1}{cd}\sum_{q=1}^{d}\left(\frac{\sum_{\ell=1}^{c}\left(BG_{qq}^{\{\ell\}}\right)'}{TSS_q} - \frac{BGSS_q}{TSS_q}\frac{\left(TSS_q\right)'}{TSS_q}\right).\end{aligned} \tag{278}$$

By using the Lagrange multiplier, we can write the Lagrangian function as

$$L(\mu, z) = RLI^2 - \sum_{i=1}^{n} \lambda_i g_i. \tag{279}$$

where $g_i = \sum_{\ell=1}^{c} \mu_{i\ell} - 1$. Taking the first order partial derivatives with respect to μ, z, $z^{\{k\}}$ and λ and putting them equal to zero, given as

$$\frac{\partial}{\partial \mu_{ik}}\left(RLI^2 - \sum_{i=1}^{n} \lambda_i g_i\right) = 0 \quad (1 \le i \le n, 1 \le k \le c), \tag{280}$$

$$\frac{\partial}{\partial z^p}\left(RLI^2 - \sum_{i=1}^{n} \lambda_i g_i\right) = 0 \quad (1 \le p \le d), \tag{281}$$

$$\frac{\partial}{\partial z^{\{k\}p}}\left(RLI^2 - \sum_{i=1}^{n} \lambda_i g_i\right) = 0 \quad (1 \le k \le c, 1 \le p \le d), \tag{282}$$

$$\frac{\partial}{\partial \lambda_i}\left(RLI^2 - \sum_{i=1}^{n} \lambda_i g_i\right) = 0 \qquad (1 \le i \le n). \tag{283}$$

Eqs. (280), (281), (282) and (283), respectively imply that

$$\frac{m\mu_{ik}^{m-1}}{cd} \sum_{q=1}^{c} \frac{1}{TSS_q} \left(\frac{BG_{qq}^{\{k\}}}{\nu^{\{k\}}} - \frac{BGSS_q}{TSS_q}(o_i^q - z^q)\right) = \lambda_i, \tag{284}$$

$$\frac{-2}{cdTSS_p}\left(\sum_{\ell=1}^{c} \nu^{\{\ell\}}(z^{\{k\}p} - z^p) - \frac{BGSS_q}{TSS_q} \nu(b^p - z^p)\right) = 0, \tag{285}$$

$$\frac{-2}{cd}\frac{\nu^{\{k\}}(z^{\{k\}p} - z^p)}{TSS_p} = 0, \tag{286}$$

$$g_i = 0. \tag{287}$$

Eq. (287) provide the original constraints and (286) imply that

$$z^{\{k\}} = z \text{ or } \nu^{\{k\}} = 0. \tag{288}$$

This means that for each cluster, the representative of the cluster coincides with the total representative or the cluster is null. To make matters worse, this implies

$$BG_{pp}^{\{k\}} = \nu^{\{k\}}(z^{\{k\}p} - z^p)^2 = 0,$$

and hence

$$BGSS_p = \sum_k BG_{pp}^{\{k\}} = 0.$$

As a result, by taking $z^{\{k\}} = z$ or $\nu^{\{k\}} = 0$ for each cluster, RLI for an optimal clustering become 0. In this case $\mu_{ik} = 0$ $(1 \le i \le n)$ for cluster with $\nu^{\{k\}} = 0$. Consequently, it turn out that parametric representatives argument does not work at all for Ratkowsky Lance Index (RLI).

4.10.2. *Barycentric Representatives*

By applying z and $z^{\{k\}}$'s as the total and within cluster barycenters, respectively, RLI can be rewritten as

$$RLI^2(\mu) = \frac{1}{cd}\sum_{p=1}^{d} \frac{\sum_{k=1}^{c} \nu^{\{k\}}(b^{\{k\}p} - b^p)^2}{\nu \mathrm{E}\left[(o_*^p - b^p)^2\right]}. \tag{289}$$

By using the Lagrange multiplier, we can write the Lagrangian function as

$$L(\mu, z) = RLI^2 - \sum_{i=1}^{n} \lambda_i g_i, \tag{290}$$

where $g_i = \sum_{\ell=1}^{c} \mu_{i\ell} - 1$. Taking the first order partial derivative with respect to μ, λ and putting them equal to zero, respectively we get similar equations to (280) and (283). Now z and $z^{\{k\}}$ are omitted because they are no longer independent variables. Furthermore, by direct calculation, the equations (280) and (283) respectively become

$$\frac{m\mu_{ik}^{m-1}}{cd} \sum_{q=1}^{d} \left(\frac{WGSS_q}{TSS_q}\frac{\left(e_q.(o_i - b)\right)^2}{TSS_q} - \frac{\left(e_q.(o_i - b^{\{k\}})\right)}{TSS_q}\right) = \lambda_i, \tag{291}$$

$$g_i = 0. \tag{292}$$

Then by solving Eq. (291) with respect to μ's, we have

$$\mu_{ik} = \left(\frac{\lambda_i cd}{m} \sum_{q=1}^{d} \frac{WGSS_q}{TSS_q} \frac{\left(e_q.(o_i - b)\right)^2}{TSS_q} \right)^{\frac{1}{m-1}} \times \left(1 - \frac{\sum_{q=1}^{d} \frac{\left(e_q.(o_i - b^{\{k\}})\right)^2}{TSS_q}}{\sum_{q=1}^{d} \frac{WGSS_q}{TSS_q} \frac{\left(e_q.(o_i - b^{\{k\}})\right)^2}{TSS_q}} \right)^{-\frac{1}{m-1}}. \tag{293}$$

Substituting them into (183), we have

$$1 = \left(\frac{\lambda_i cd}{m} \sum_{q=1}^{d} \frac{WGSS_q}{TSS_q} \frac{\left(e_q.(o_i - b)\right)^2}{TSS_q} \right)^{\frac{1}{m-1}} \times \sum_{k=1}^{c} \left(1 - \frac{\sum_{q=1}^{d} \frac{\left(e_q.(o_i - b^{\{k\}})\right)^2}{TSS_q}}{\sum_{q=1}^{d} \frac{WGSS_q}{TSS_q} \frac{\left(e_q.(o_i - b)\right)^2}{TSS_q}} \right)^{-\frac{1}{m-1}}. \tag{294}$$

Hence, we get the membership degree μ_{ik} as

$$\mu_{ik} = \frac{\left(1 - \frac{\sum_{q=1}^{d} \frac{\left(e_q.(o_i - b^{\{k\}})\right)^2}{TSS_q}}{\sum_{q=1}^{d} \frac{WGSS_q}{TSS_q} \frac{\left(e_q.(o_i - b^{\{k\}})\right)^2}{TSS_q}} \right)^{-\frac{1}{m-1}}}{\sum_{\ell=1}^{c} \left(1 - \frac{\sum_{q=1}^{d} \frac{\left(e_q.(o_i - b^{\{\ell\}})\right)^2}{TSS_q}}{\sum_{q=1}^{d} \frac{WGSS_q}{TSS_q} \frac{\left(e_q.(o_i - b^{\{\ell\}})\right)^2}{TSS_q}} \right)^{-\frac{1}{m-1}}}. \tag{295}$$

In this index, the cluster center updating equation is calculated as the barycenter b and the membership degree updating equations is calculated as (295) .

4.11. *Optimization of Scott-Symons Index (SSI)*

Regarding the parametric within cluster center $z^{\{k\}}$'s as the representative point of cluster, the FCM problem is to minimize the objective function of SSI, given as

$$SSI = \sum_{k=1}^{c} \nu \log\det \left(\frac{\varpi^{\{k\}} \left[(o - z^{\{k\}})\ {}^t(o - z^{\{k\}}) \right]}{\nu^{\{k\}}} \right).$$

By using the Lagrange multiplier, we can write the Lagrangian function as

$$L(\mu, z, \lambda) = SSI - \sum_{i=1}^{n} \lambda_i g_i, \tag{296}$$

where $g_i = \sum_{\ell=1}^{c} \mu_{i\ell} - 1$. Then taking the first order partial derivatives with respect to $\mu, z^{\{k\}}$ and λ and putting them equal to zero, respectively we get

$$\frac{\partial}{\partial \mu_{ik}} \left(SSI - \sum_{i=1}^{n} \lambda_i g_i \right) = 0 \qquad (1 \le i \le n, 1 \le k \le c), \tag{297}$$

$$\frac{\partial}{\partial z^{\{k\}p}} \left(SSI - \sum_{j=1}^{n} \lambda_j g_j \right) = 0 \qquad (1 \le k \le c, 1 \le p \le d), \tag{298}$$

$$\frac{\partial}{\partial \lambda} \left(SSI - \sum_{j=1}^{n} \lambda_j g_j \right) = 0 \qquad (1 \le i \le n). \tag{299}$$

Eqs. (297), Eq. 298), (299) respectively imply

$$m\mu_{ik}^{m-1} \left(\log\det \frac{WG^{\{k\}}}{\nu^{\{k\}}} - d + \|o_i - z^{\{k\}}\|^2_{(WG^{\{k\}}/\nu^{\{k\}})^{-1}} \right) = \lambda_i, \tag{300}$$

$$-2\nu^{\{k\}t} e_p (WG^{\{k\}}/\nu^{\{k\}})^{-1} (b^{\{k\}} - z^{\{k\}}) = 0. \tag{301}$$

$$g_i = 0. \tag{302}$$

Eq. (302) provide the original constraints and as usual (301) cause

$$z^{\{k\}} = b^{\{k\}} = \mathrm{E}^{\{k\}}[o]. \tag{303}$$

And hence the Eq. (300) are rewritten as

$$m\mu_{ik}^{m-1} \left(\log\det \Sigma^{\{k\}} - d + \|o_i - b^{\{k\}}\|^2_{(\Sigma^{\{k\}})^{-1}} = \lambda_i. \right. \tag{304}$$

Thus μ' become like this

$$\mu_{ik} = \left(\frac{\lambda_i}{m}\right)^{\frac{1}{m-1}} \Bigg(\log\det \Sigma^{\{k\}} - d + \|o_i - b^{\{k\}}\|^2_{(\Sigma^{\{k\}})^{-1}}\Bigg)^{-\frac{1}{m-1}} \tag{305}$$

then substitute them into (183), we get

$$1 = \left(\frac{\lambda_i}{m}\right)^{\frac{1}{m-1}} \sum_{\ell=1}^{c} \Bigg(\log\det \Sigma^{\{k\}} - d + \|o_i - b^{\{k\}}\|^2_{(\Sigma^{\{k\}})^{-1}}\Bigg)^{-\frac{1}{m-1}}. \tag{306}$$

And we obtain the updating rule for membership degree μ_{ik} as

$$\mu_{ik} = \frac{\left(\log\det \Sigma^{\{k\}} - d + \|o_i - b^{\{k\}}\|^2_{(\Sigma^{\{k\}})^{-1}}\right)^{-\frac{1}{m-1}}}{\sum_{\ell=1}^{c}\left(\log\det \Sigma^{\{\ell\}} - d + \|o_i - b^{\{\ell\}}\|^2_{(\Sigma^{\{\ell\}})^{-1}}\right)^{-\frac{1}{m-1}}}. \tag{307}$$

The within cluster center updating equation is in calculated as (303) and the the membership degree updating equation is calculated as (307).

The Barycentric representative in this index is similar discussion to the previous index with the parametric total center z and the parametric within cluster center $z^{\{k\}}$ are equal to the total barycenter b and the within cluster barycenter $b^{\{k\}}$ respectively.

4.12. *Optimization of Trace W Index (TrWI)*

Regarding the parametric within cluster center $z^{\{k\}}$'s as the representative point of cluster, the FCM problem is to minimize the objective function of TWI, given as

$$TWI = \sum_{k=1}^{c} \varpi^{\{k\}}\left[\left\|(o - z^{\{k\}})\right\|^2\right]. \tag{308}$$

By using the Lagrange multiplier, we can write the Lagrangian function as

$$L(\mu, z, \lambda) = TWI - \sum_{i=1}^{n} \lambda_i g_i, \tag{309}$$

where $g_i = \sum_{\ell=1}^{c} \mu_{i\ell} - 1$. Then taking the first order partial derivatives with respect to $\mu, z^{\{k\}}$ and λ and putting them equal to zero, respectively we get

$$\frac{\partial}{\partial \mu_{ik}}\left(TWI - \sum_{i=1}^{n}\lambda_i g_i\right) = 0 \quad (1 \le i \le n, 1 \le k \le c), \tag{310}$$

$$\frac{\partial}{\partial z^{\{k\}p}}\left(TWI - \sum_{i=1}^{n}\lambda_i g_i\right) = 0 \quad (1 \le k \le c, 1 \le p \le d), \tag{311}$$

$$\frac{\partial}{\partial \lambda}\left(TWI - \sum_{i=1}^{n}\lambda_i g_i\right) = 0 \qquad (1 \le i \le n). \tag{312}$$

Eqs. (310), (311), (312) respectively imply

$$(\partial_{ik}\nu^{\{k\}})\|o_i - z^{\{k\}}||^2 = \lambda_i, \tag{313}$$

$$-2\nu^{\{k\}t} e_P (b^{\{k\}} - z^{\{k\}}) = 0, \tag{314}$$

$$g_i = 0. \tag{315}$$

Eq. (315) provide the original constraints and as usual (314) cause

$$z^{\{k\}} = b^{\{k\}} = \mathrm{E}^{\{k\}}[o]. \tag{316}$$

And hence (313) are rewritten as

$$m\mu_{ik}^{m-1}\|o_i - b^{\{k\}}||^2 = \lambda_i. \tag{317}$$

Thus μ's become

$$\mu_{ik} = \left(\frac{\lambda_i}{m}\right)^{-\frac{1}{m-1}} \|o_i - b^{\{k\}}\|^{-\frac{2}{m-1}}. \tag{318}$$

Then substitute them into (183), we have

$$1 = \left(\frac{\lambda_i}{m}\right)^{\frac{1}{m-1}} \sum_{\ell=1}^{c} \|o_i - b^{\{k\}}\|^{-\frac{2}{m-1}}. \tag{319}$$

Hence we get the membership degree as

$$\mu_{ik} = \frac{\|b^{\{k\}} - o_i\|^{-\frac{2}{m-1}}}{\sum_{\ell=1}^{c}\|b^{\{\ell\}} - o_i\|^{-\frac{2}{m-1}}}. \tag{320}$$

The within cluster center updating equation is calculated as (316) and the membership degree updating equation is calculated as (320).

The Barycentric representative in this index is similar discussion to the previous index.

4.13. *Optimization of Trace WiB Index (TrWiBI)*

Regarding the parametric total center z and the parametric within cluster centers $z^{\{k\}}$ as the representative points of all data and data in clusters, respectively, the FCM problem is to minimize the objective function of TrWiBI, given as

$$TrWiBI = \sum_{\ell=1}^{c} \nu^{\{\ell\}} \|z^{\{\ell\}} - z\|^2_{WG^{-1}}.$$

By using the Lagrange multiplier, we can write the Lagrangian function as

$$L(\mu, z, \lambda) = TrWiBI - \sum_{i=1}^{n} \lambda_i g_i \tag{321}$$

where $g_i = \sum_{\ell=1}^{c} \mu_{i\ell} - 1$. Then taking the first order partial derivatives with respect to $\mu, z, z^{\{k\}}$ and λ and putting them equal to zero, respectively we get

$$\frac{\partial}{\partial \mu_{ik}} \left(TrWiBI - \sum_{i=1}^{n} \lambda_i g_i \right) = 0 \quad (1 \le i \le n, 1 \le k \le c), \tag{322}$$

$$\frac{\partial}{\partial z^p} \left(TrWiBI - \sum_{i=1}^{n} \lambda_i g_i \right) = 0 \quad (1 \le p \le d), \tag{323}$$

$$\frac{\partial}{\partial z^{\{k\}p}} \left(TrWiBI - \sum_{i=1}^{n} \lambda_i g_i \right) = 0 \quad (1 \le k \le c, 1 \le p \le d), \tag{324}$$

$$\frac{\partial}{\partial \lambda} \left(TrWiBI - \sum_{i=1}^{n} \lambda_i g_i \right) = 0 \qquad (1 \le i \le n). \tag{325}$$

Eqs. (322), (323), (324) and (325) respectively imply that

$$m\mu_{ik}^{m-1} \left(\|z^{\{k\}} - z\|^2_{WG^{-1}} - \sum_{\ell=1}^{c} \nu^{\{\ell\}} \left\langle z^{\{k\}} - z, o_i - z^{\{k\}} \right\rangle^2_{WG^{-1}} \right) = \lambda_i, \tag{326}$$

$$2 \left\langle e_p, \nu z - \sum_{\ell=1}^{c} \nu^{\{\ell\}} z^{\{\ell\}} \right\rangle_{WG^{-1}} = 0, \tag{327}$$

$$2\nu^{\{k\}} \left\langle e_p, (z^{\{k\}} - z) - BGWG^{-1}(z^{\{k\}} - b^{\{k\}}) \right\rangle_{WG^{-1}} = 0, \tag{328}$$

$$g_i = 0. \tag{329}$$

Eq. (329) provide the original constraints and (327) and (328) cause

$$z = \sum_{k=1}^{c} \frac{\nu^{\{k\}}}{\nu} z^{\{\ell\}} \tag{330}$$

$$z^{\{k\}} - z = BGWG^{-1}(z^{\{k\}} - b^{\{k\}}). \tag{331}$$

Hence, similar as usual, we obtain the center must be a barycenter, given as

$$z = b = \mathrm{E}[o]. \tag{332}$$

$$z^{\{k\}} = b^{\{k\}} = \mathrm{E}^{\{k\}}[o]. \tag{333}$$

Eq. (326) can be rewritten as

$$m\mu_{ik}^{m-1} TrWiBI \big(\|o_i - z\|^2_{T^{-1}} - \|o_i - z^{\{k\}}\|^2_{WG^{-1}} \big) = \lambda_i, \tag{334}$$

thus μ_{ik} become like this

$$\mu_{ik} = \left(\frac{c\lambda_i}{mTrWiBI} \right)^{\frac{1}{m-1}} \big(\|o_i - z\|^2_{T^{-1}} - \|o_i - z^{\{k\}}\|^2_{WG^{-1}} \big)^{\frac{1}{m-1}}. \tag{335}$$

Then substituting them into Eq. (183), we have

$$1 = \left(\frac{c\lambda_i}{mTrWiBI} \right)^{\frac{1}{m-1}} \sum_{\ell=1}^{c} \big(\|o_i - z\|^2_{T^{-1}} - \|o_i - z^{\{\ell\}}\|^2_{WG^{-1}} \big)^{\frac{1}{m-1}}. \tag{336}$$

Hence we get the equations for membership degree updating rule as

$$\mu_{ik} = \frac{\big(\|o_i - b\|^2_{T^{-1}} - \|o_i - b^{\{k\}}\|^2_{WG^{-1}} \big)^{-\frac{1}{m-1}}}{\sum_{\ell=1}^{c} \big(\|o_i - b\|^2_{T^{-1}} - \|o_i - b^{\{\ell\}}\|^2_{WG^{-1}} \big)^{-\frac{1}{m-1}}} \tag{337}$$

The cluster center updating equation is calculated as (333) and the membership degree updating equation is calculated as (337).

The Barycentric representative in this index is similar discussion to the previous index.

5. Experiment Analysis

In this section, the proposed methods are employed for COVID 19 data in Indonesia which consist of 34 provinces and 3 variables such as confirmed cases, recovered cases and death cases. This data is provided by Indonesian Health Ministry as per March 2nd, 2020 until September 7th, 2021 (https://covid19.go.id/id) that the result of COVID 19 in Indonesia shows in tabel below.

Table 1: Statistic of Covid-19 Case in Indonesia.

Cases	Min	Mean	Max
Confirmed	6.189	84.203	649.309
Recovered	561	54.962	543.801
Death	129	1.932	13.806

Based on the algorithm of FCM described in the previous section, the first step is to initialize (or) the initialization of the membership degree $\mu_{ik}(0)$ randomly. In this section, the initialization of $\mu_{ik}(0)$ are same for the objective function J and all the analytic indices to know the behaviors of indices in the same initial value. And then we update the membership degree based on the updating rule equations of membership degree depending on the objective function J and each of the analytic indices. Hence, we have a set of sequences of membership degrees such that each of them yields the value of the l objective function or individual analytic indices. Fig.2 shows our design of experiment.

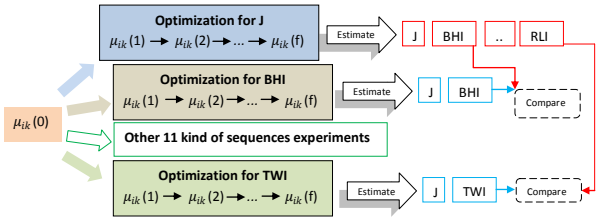

Fig. 2: Design of Experiment

The top blue box indicates a sequence for optimizing process for the objective function J. Fig.3 shows the results of applying the values membership degree at each step of this data sequence to the objective function and other 13 analytic indices. In this case, we get the result of J and other 13 indices that can be seen in Fig.3.

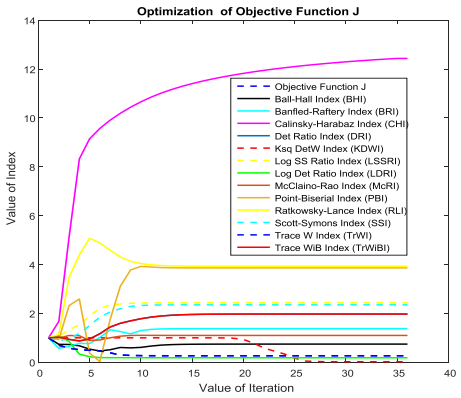

Fig. 3: Optimization for J

The graphs plotted in Fig.3 show that most of the indicators do not work well in clustering, as the optimization process is for the objective function. Only $KDWI$ and $LDRI$ have good performance. For the objective function J and $TrWI$ are shown by the same line (blue dotted line) because the formulas of both are similar, hence the results are also similar. We can see also that almost all indices have the fluctuation in each iteration thus they are very difficult to determine whether the final iteration of the indices produce an optimal cluster or not. Note that for all indices (include the objective function), the values start from 1 because of normalization.

Similarly, in our experiment, we optimize the 13 clustering indices mark out on the Fig.2 (grey box until green box) with the goal to obtain the optimal cluster directly. Main quantities that characterize the optimal cluster are the density or the within-cluster and the separability or the between-cluster. Some indices are to be maximized and the others are minimized based on their functions to obtain the optimal cluster. Fig.4 shows increasing and decreasing trends for the optimization of the objective function for each indicator. The plots in each of the boxes in Fig.4 are the comparison between the behavior of the index obtained from the optimization for J (red lines) and obtained from optimization for the index itself (dotted blue line).

For more details, we show the comparisons in the Table 2. There are the values for each index in the final iteration. The indices consist of

(a) Ball-Hall Index (BHI). The minimum value of this index indicates the optimal clusters.
(b) Banfled-Raftery Index (BRI). The minimum value of this index indicates the optimal clusters.
(c) Chalinsky-Harabaz Index (CHI). The maxi-

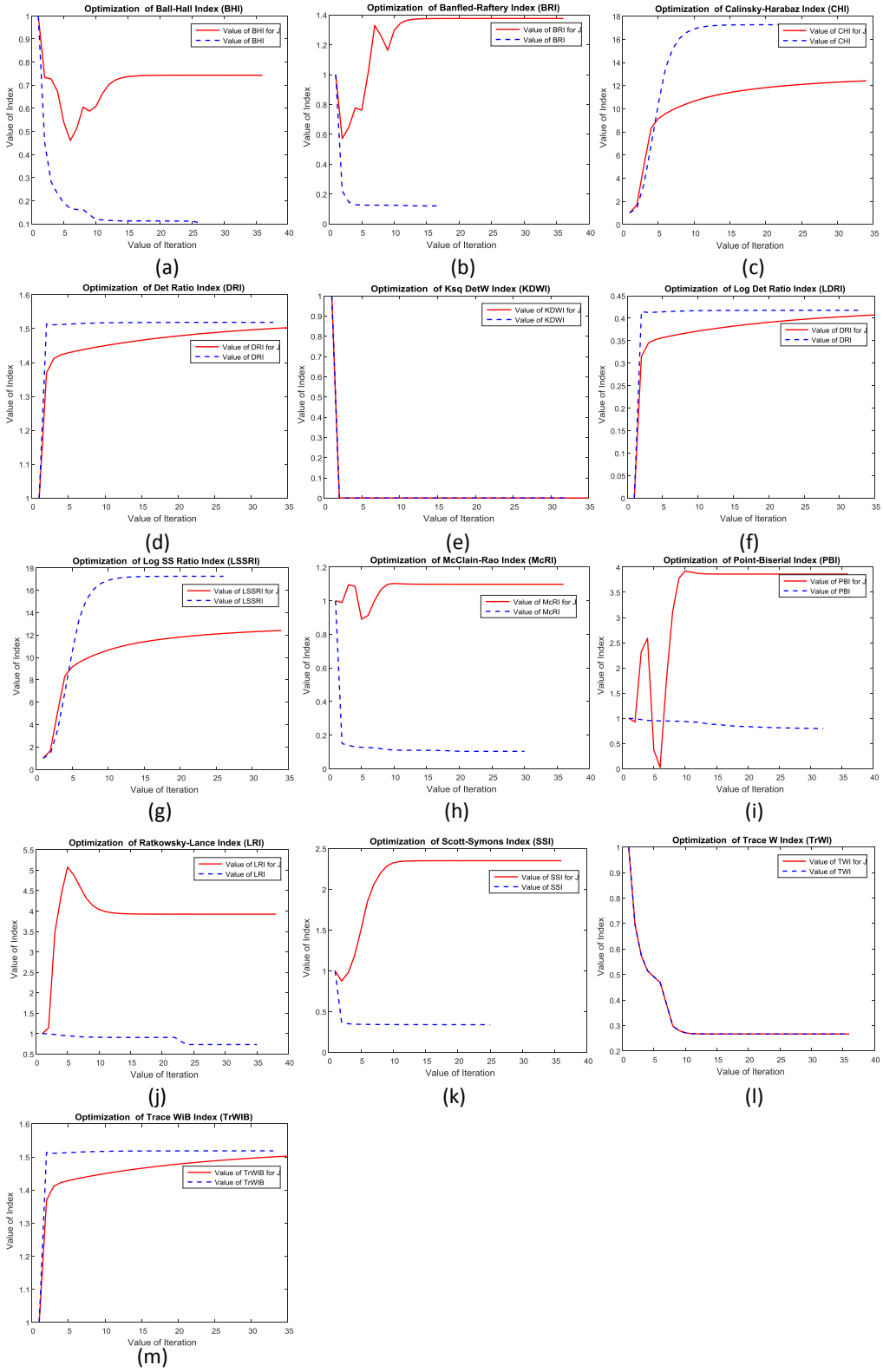

Fig. 4: Plots of Optimization for All Indices. (a) Ball-Hall Index (BHI); (b) Banfled-Raftery Index (BRI); (c) Calinsky-Harabaz Index (CHI); (d) Det Ratio Index (DRI); (e) Ksq DetW Index (KDWI); (f) Log Det Ratio Index (LDRI); (g) Log SS Ratio Index (LSSRI); (h) McCLain-Rao Index (McRI); (i) Point-Biserial Index (PBI); (j) Ratkowsky-Lance Index (RLI); (k) Scott-Symons Index (SSI); (l) Trace W Index (TrWI); (m) Trace WiB Index (TrWiBI).

Table 2: Result of Comparison of Indices

Indices	BHI	BRI	CHI	DRI	KDWI	LDRI	LSSRI	McRI	PBI	RLI	SSI	TrWI	TrWiBI
Optimization for J	1.06	302.8	83.11	15.85	0.02	0.40	12.41	10.97	38.61	1.61	23.50	2.34	15.85
Optimization for Indices	0.21	27.27	115.58	16.15	0.002	0.41	17.25	0.699	4.75	0.731	16.15	2.34	16.15

mum value is the better performance.

(d) Det Ratio Index (DRI). The maximum value of this index is the better performance.

(e) Ksq DetW Index (KDWI). The minimum value of the index indicates the optimal clusters.

(f) Log Det Ratio Index (LDRI). The maximum value indicates the optimal clusters.

(g) Log SS Ratio Index (LSSRI) . The maximum value indicates the optimal clusters..

(h) McCLain-Rao Index (McRI). The minimum value of this index is the better performance.

(i) Point-Biserial Index (PBI). The minimum value of this index indicates the optimal clusters.

(j) Ratkowsky-Lance Index (RLI). The minimum value is the better performance.

(k) Scott-Symons Index (SSI). The minimum value of this index indicates the optimal clusters.

(l) Trace W Index (TrWI). The minimum value indicates the optimal clusters.

(m) Trace WiB Index (TrWiBI). The maximum value indicates the optimal clusters.

6. Conclusion

In this article, we propose 2 methods namely fuzzification of clustering indices and optimization of clustering indices. Fuzzification can be performed from the crisp clustering indices, in a way that includes themselves as special cases. The process is achieved by converting the crisp quantities {0,1} to the fuzzy quantities [0,1]. A significant benefit of fuzzy clustering is that the membership degrees allow the optimization problems to be treated as continuous, whereas the ones in the crisp case are discrete, making it generally easier to solve.

Clustering indices is used to evaluate the quality of clustering results. By optimizing the fuzzified clustering indices, various new fuzzy clustering methods can be proposed. In this article, we only choose the analytical indices to be optimized by the gradient method. The results show that by optimizing the clustering indices, we can reach the optimal clustering directly.

References

1. D. R. Sope and M. Fujio, On a fuzzification of clustering indices, in *Proc. Fuzzy System Symposium of the Japan Intelligent Information Fuzzy Society* **35** (2002) 443–448.
2. B. Desgraupes, Clustering Indices, (2016).
3. J. Bezdek, R. Ehrlich and W. Full, FCM: The fuzzy c-means clustering algorithm, *Computers and Geosciences* **10**(2) (1984) 191–203.
4. K. L. Wu, Analysis of parameter selections for fuzzy c-mean, *Pattern Recognition* **45**(1) (2012) 407–415.
5. N. R. Pal and J. C. Bezdek, On cluster validity for fuzzy c-means model, *IEEE Trans. on Fuzzy Syst.* **3**(3) (1995) 370–379.
6. J. Yu, Q. Cheng and H. Huang, Analysis of the weighting exponent in the FCM, *IEEE Trans. Syst. Man and Cybernetic Part B (Cybernetic)* **34**(1) (2004) 634–639.
7. J. Pei, X. Yang, X. Gao and W. Xie, Weighting exponent m in fuzzy c-means (FCM) clustering algorithm, *Proc. of SPIE - The International Society for Optical Engineering* (2001), doi: 10.1117/12.441637.
8. G. Xin, A study of weighting exponent m in a fuzzy c means algorithm, (2000).
9. J. R. Nolan, A Prototype application of fuzzy logic and expert systems in education assessment, *IAAI-98 Proc.* (1998) 1134–1139.
10. M. J. Fadili, S. Ruan, D. Bloyet and B. Mazoyer, On the number of clusters and the fuzziness index for unsupervised FCA application to BOLD fMRI time series, *Medical Image Analysis* **5**(1) (2001) 55–67, doi: 10.1016/s1361-8415(00)00035-9.
11. L. A. Goodman and W. H. Kruskal, Measures of associations for cross-validations, *Journal of the American Stat. Assoc.* **49** (1954) 732–764.
12. M. K. Pakhira, S. Bandyopadhyay and U. Maulik, Validity index for crisp and fuzzy clusters, *Pattern Recognition* **37**(3) (2004) 487–501, `https://www.sciencedirect.com/science/article/pii/S0031320303002838`
13. L. Hubert and J. Schultz, Quadratic assignment as a general data analysis strategy, *British Journal of Mathematical and Statistical Psychology* **29**(2) (1976) 190–241.
14. P. S. Bradley, U. M. Fayyad and C. A. Reina, Scaling EM (Expectation-Maximization) clustering to large

databases, (2000).
15. H. Hruschka, Market definition and segmentation using fuzzy clustering methods, *International Journal of Research in Marketing* **3**(2) (1986) 117–134, `https://www.sciencedirect.com/science/article/pii/0167811686900157`.
16. B. Chérief-Abdellatif and P. Alquier, Consistency of variational bayes inference for estimation and model selection in mixtures, *Electronic Journal of Statistics* **12**(2) (2018) 2995–3035.
17. D. R. Brillinger and H. K. Preisler, Maximum likelihood estimation in a latent variable problem, *Studies in Econometrics, Time Series, and Multivariate Statistics* (Academic Press, 1983) 31–65.
18. I. B. Vapnyarskii, Lagrange multipliers, *Encyclopedia of Mathematics* **978** (Springer, 2001) 1–55.
19. J. T. Tou and R. C. Gonzales, Pattern recognition principles [by] Julius T. Tou [and] Rafael C. Gonzalez, *Addison-Wesley Pub. Co Reading, Mass* (1974).
20. D. P. Bertsekas, Constrained optimization and Lagrange multiplier methods, (Academic Press, 2014).