Capacitated Clustering via Majorization-Minimization and Collaborative Neurodynamic Optimization

Hongzong Li and Jun Wang, *Life Fellow, IEEE*

***Abstract*— This paper addresses capacitated clustering based on majorization-minimization and collaborative neurodynamic optimization (CNO). Capacitated clustering is formulated as a combinatorial optimization problem. Its objective function consists of fractional terms with intra-cluster similarities in their numerators and cluster cardinalities in their denominators as normalized cluster compactness measures. To obviate the difficulty in optimizing the objective function with factional terms, the combinatorial optimization problem is reformulated as an iteratively reweighted quadratic unconstrained binary optimization problem with a surrogate function and a penalty function in a majorization-minimization framework. A clustering algorithm is developed based on CNO for solving the reformulated problem. It employs multiple Boltzmann machines operating concurrently for local searches and a particle swarm optimization rule for repositioning neuronal states upon their local convergence. Experimental results on ten benchmark datasets are elaborated to demonstrate the superior clustering performance of the proposed approaches against seven baseline algorithms in terms of 21 internal cluster validity criteria.**

***Index Terms*— Capacitated clustering, collaborative neurodynamic optimization (CNO), iteratively reweighted optimization, majorization-minimization, quadratic unconstrained binary optimization.**

I. Introduction

CLUSTERING is one of the most fundamental means of data processing. It is to find a set of clusters in unlabeled data. In many applications, some prior information on the clusters is available in the form of constraints or partially labeled data. As a basic form of semi-supervised learning, constrained clustering incorporates available prior information/knowledge into clustering. Clustering performance can be boosted by leveraging the constraints that confine the search space. Several types of constraints are utilized in constrained clustering: cluster-level constraints, such as cardinality constraints [1], [2], [3], [4], [5], [6], [7], [8] and capacity constraints [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19], [20], [21], [22], [23], [24], [25], [26], instance-level constraints [27], such as must-link and cannot-link constraints [28], and rank constraint [29].

Capacitated clustering is an important type of constrained clustering. It arises in many applications, including vehicle routing [30], [31], [32], very large-scale integrated circuit design [33], mail delivery [34], sibling reconstruction [35], mobility network handover minimization [20], [36], facility layout [37], [38], reviewer groups construction [39], wireless and wired network integration [40], and stock market index tracking [41].

In the literature, capacitated clustering is commonly formulated as a combinatorial optimization problem [9]. In the existing problem formulation, the objective function is composed of multiple functional terms, and each term sums the dissimilarity coefficients (e.g., distances) between every pair of data in a cluster to quantify the within-cluster compactness [9], [11], [12], [16], [17], [19], [20], [21], [23], [24], [25], [26]. Existing capacitated clustering methods can be classified into two major categories: mathematical programming methods and heuristic or meta-heuristic methods. Specifically, mathematical programming methods include the branch and cut method [19], the Lagrangian relaxation method [17], etc. In view that most combinatorial optimization problems are NP-hard [42], heuristic and meta-heuristic methods are widely used, including scatter search [12], tabu search (TS) [18], variable neighborhood search [25], neighborhood decomposition-driven variable neighborhood search (NDVNS) [26], iterated variable neighborhood search (IVNS) [21], greedy random adaptive search procedure and adaptive memory programming [11], reactive greedy randomized adaptive search procedure with path relinking [16], TS with variable neighborhood search [20], greedy randomized adaptive search procedure with TS [20], greedy randomized adaptive search procedure with variable neighborhood search [20], general variable neighborhood search (GVNS) [25], skewed GVNS (SGVNS) [25], etc. Although heuristic and meta-heuristic methods can often find good solutions, the solution quality varies without guaranteeing global optimality.

Manuscript received 21 July 2022; accepted 4 October 2022. This work was supported in part by the Research Grants Council of the Hong Kong Special Administrative Region of China under Grant 11202019 and Grant11203721 and in part by the Laboratory for AI-Powered Financial Technologies. *(Corresponding author: Jun Wang.)*

Hongzong Li is with the Department of Computer Science, City University of Hong Kong, Hong Kong (e-mail: hongzli2-c@my.cityu.edu.hk).

Jun Wang is with the Department of Computer Science and the School of Data Science, City University of Hong Kong, Hong Kong (e-mail: jwang.cs@cityu.edu.hk).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TNNLS.2022.3212593.

Digital Object Identifier 10.1109/TNNLS.2022.3212593

Over the last few decades, neurodynamic optimization approaches emerged as computationally intelligent optimizers based on recurrent neural networks for solving various optimization problems, such as combinatorial optimization problems [43]. It is very difficult, if not impossible, for an individual recurrent neural network to solve a combinatorial optimization problem, as it may easily get stuck in a local minimum. As such, more than one neurodynamic model is needed to work collaboratively. In recent years, a hybrid intelligence framework called collaborative neurodynamic optimization (CNO) has been developed based on multiple neurodynamic models coordinated by using meta-heuristics. The multiple neurodynamic models operate concurrently for scattered local searches, and their initial states are reinitialized repeatedly by using a meta-heuristic rule for repositioning the searches. It integrates the scatter search capability of multiple neurodynamic models together with the global search capability of meta-heuristics. It is proven that CNO approaches are almost surely convergent (i.e., with probability one) to the global optimal solutions of optimization problems [44], [45], [46], [47], [48].

In this paper, we first propose an objective function with multiple functional terms normalized by the cluster sizes to measure the within-cluster compactness. Next, we formulate capacitated clustering as an iteratively reweighted quadratic unconstrained binary optimization problem with a surrogate function and a penalty function in a majorization-minimization framework. We develop a CNO-driven capacitated clustering algorithm based on a population of Boltzmann machines (BMs) with momentum terms for solving the reformulated problem. The novelties and contributions of this work are summarized as follows.

1) The reformulated objective function with the total within-cluster dissimilarity of every cluster normalized by its cluster cardinality is able to characterize the cluster compactness naturally, independent of cluster sizes.
2) The iteratively reweighted quadratic unconstrained binary optimization problem reformulation with a surrogate function enables avoiding the direct use of the reformulated objective function with factional terms while keeping their cluster normalization effect.
3) The CNO-driven clustering algorithm with a population of BMs operating in parallel fully utilizes the hill-climbing capability of BMs in scattered searches.
4) Extensive experimental results on ten benchmark datasets show that the CNO-driven clustering algorithm statistically outperforms seven prevailing baselines in most internal cluster validity indices.

The remaining paper is arranged as follows. In Section II, background knowledge on neurodynamic optimization to be used in Section IV is introduced. In Section III, capacitated clustering is reformulated as an iteratively reweighted quadratic unconstrained binary optimization problem. In Section IV, a CNO-driven capacitated clustering algorithm is delineated. In Section V, experimental results are discussed in detail. In Section VI, concluding remarks are given.

II. Background Knowledge

A. Neurodynamic Optimization Models

1) Discrete Hopfield Network: The discrete Hopfield network (DHN) is an exemplar of recurrent neural networks characterized by its binary or bipolar states and hard-limiter activation function operating in discrete time [49]

$$\begin{cases} u(t+1) = Wx(t) - \theta \\ x(t) = \sigma(u(t)) \end{cases} \tag{1}$$

where $u \in \Re^n$ is the net-input vector, $x \in \Re^n$ is the state vector, $W \in \mathbb{R}^{n\times n}$ is the connection weight matrix, $\theta \in \mathbb{R}^n$ is the threshold vector, and $\sigma(\cdot)$ is a vector-valued hard-limiter activation function defined elementwise as follows:

$$\sigma(u_i) = \begin{cases} 0, & \text{if } u_i(t) \le 0 \\ 1, & \text{if } u_i(t) > 0. \end{cases}$$

It is shown in [49] that the DHN in (1) is completely stable; i.e., the state of the DHN is convergent to an equilibrium from any initial state, and if the connection weight matrix is symmetric (i.e., $W = W^T$), the main diagonal elements of W are zero (i.e., $w_{ii} = 0, \forall i$), and the activation takes place asynchronously. It is also shown in [49] that the DHN is convergent to a local minimum of the following binary optimization problem:

$$\min -\frac{1}{2}x^T W x + \theta^T x \quad \text{s.t. } x \in \{0,1\}^n. \tag{2}$$

It means that an equilibrium $\bar{x}$ is a local optimum of the optimization problem earlier. As the right-hand side of (1) is equal to the negative gradient of the objective function to be minimized, the DHN neurodynamics forms a gradient flow moving among vertices of the unit hypercube coordinatewise. Note that the binary states of the DHN depend exclusively on the sign of the negative gradient of the objective function without any historical effect.

If W in the quadratic term of (2) is asymmetric in a given problem, its symmetry can be equivalently realized by replacing it with $(W + W^T)/2$. The zero diagonal elements of W can equivalently realized by adding a linear term $\mathrm{diag}(w_{11}, \ldots, w_{nn})x$, in view that $x_i^2 = x_i$ for $i = 1, 2, \ldots, n$ due to the binary nature of the state variable $x_i \in \{0, 1\}$.

As a variant of the DHN, a DHN with a momentum term (DHNm) [50] is developed as follows:

$$\begin{cases} u(t+1) = u(t) + Wx(t) - \theta \\ x(t) = \sigma(u(t)) \end{cases} \tag{3}$$

where $u(0) = 0$. With the addition of the momentum term $u(t)$ in the DHN dynamic equation, the DHNm in (3) takes its historical effect into account and enriches its dynamic behaviors. It is shown that all neuronal states in the DHNm in (3) can be activated synchronously and are convergent to local or near optima [51], [52]. The DHNm is widely used in many applications, such as graph planarization [50], independent set maximization [53], checkerboard tiling [54], data sorting [55], map coloring [51], string search [56], channel assignment in cellular radio networks [57], magnetic resonance image segmentation [58], bipartite subgraph matching [59], broadcast scheduling [60], facility layout [61], via minimization [62], programmable logic array folding [63], microcode optimization [64], etc.

2) Boltzmann Machine: The BM [65] is a stochastic version of the DHN. For searching and optimization, it is a parallel realization of simulated annealing [66] minimizing (2). Unlike the Hopfield network, the ith neuron in a BM is activated according to the probability defined as follows:

$$\begin{cases} p(x_i(t)=1) = \dfrac{1}{1+\exp\left(-\frac{u_i(t)}{T(t)}\right)} \\ p(x_i(t)=0) = 1 - p(x_i(t)=1) \end{cases} \tag{4}$$

where u_i is defined as in (1), T is the temperature parameter decreasing over time. An exponential multiplicative cooling schedule of the temperature is defined in [67]

$$T(t) = T_0\alpha^t \tag{5}$$

where T_0 is the initial temperature and $\alpha \in (0,1)$ is a cooling rate parameter. It is known that a BM with a sufficiently large initial temperature and sufficiently long cooling schedule is almost surely convergent to a global optimal solution of a given combinatorial optimization problem [68], [69].

The BM is applicable to a wide range of combinatorial optimization problems, including the independent set problem [68], the max-cut problem [68], the graph coloring problem [68], the traveling salesman problem [69], the matching problem [66], and the graph partitioning problem [66].

B. Collaborative Neurodynamic Optimization

CNO approaches are developed for solving various complex optimization problems, such as distributed optimization [70], [71], global optimization [44], [45], [46], [47], multiobjective optimization [71], [72], and combinatorial and mixed-integer optimization [46], [48]. They inherit the almost-sure convergence property proven theoretically in [46]. CNO is customized as computationally intelligent optimizers in various applications, such as nonnegative matrix factorization [73], vehicle-task assignment [74], [75], stock portfolio selection [76], spiking neural network regularization [77], and hash bit selection [78].

The neurodynamic models used in existing CNO approaches include projection neural networks [72], [73], [74], [76], [79] and DHNs (1) [75], [78]. Almost all of the CNO algorithms use a particle swarm optimization rule to reposition the neural searches. The standard rule is defined in [80]

$$\begin{aligned} v_i(t+1) &= c_0 v_i(t) + c_1 r_1\left(x_i^*(t) - x_i(t)\right) \\ &\quad + c_2 r_2\left(x^*(t) - x_i(t)\right) \\ &\text{if } (r_3 < S(v_{\text{id}}(t))), \quad \text{then } x_{\text{id}}(t) = 1, \text{ else } x_{\text{id}}(t) = 0 \end{aligned} \tag{6}$$

where $x_i(t)$ is the position of the ith particle at the tth iteration, $v_i(t)$ is the velocity of the ith particle, $x_i^*(t)$ is the present best position in terms of a given objective function for the ith particle, $x^*(t)$ is the best position of the population, c_0 is a positive inertia parameter, c_1 and c_2 are two positive weighting parameters, and r_1 and r_2 and r_3 are random values in [0, 1], $S(\cdot)$ is a sigmoid limiting transformation.

III. Problem Formulations

Suppose that n data vectors with m features $\underline{v}_i \in \Re^m$ are to be clustered into p mutually exclusive clusters. The capacitated clustering problem is commonly formulated as follows [9], [11], [12], [16], [17], [19], [20], [21], [23], [24], [25], [26]:

$$\begin{aligned} \min_x \quad & \sum_{l=1}^{p}\sum_{i=1}^{n}\sum_{j<i} d_{ij}x_{il}x_{jl} \\ \text{s.t.} \quad & \sum_{l=1}^{p} x_{il} = 1, \quad i = 1,2,\ldots,n \\ & \sum_{i=1}^{n} a_{il}x_{il} \le b_l, \quad l = 1,2,\ldots,p \\ & x_{il} \in \{0,1\}, \quad i = 1,2,\ldots,n;\ l = 1,2,\ldots,p \end{aligned} \tag{7}$$

where x_{il} is the binary decision variable defined as $x_{il} = 1$ if datum i is assigned to cluster l or $x_{il} = 0$, otherwise, d_{ij} is the dissimilarity coefficient (e.g., distance) between samples i and j, a_{il} is a capacity coefficient for datum i and cluster l, and b_l is the capacity of cluster l. It contains np binary decision variables and $n+p$ linear constraints. The first set of constraints ensures that each datum belongs to exactly one cluster. The second set contains capacity constraints that restrict the cluster size.

Different from k-means and k-medoid clustering, the objective function in the quadratic assignment problem formulation is centerless without the need for selecting cluster representatives (e.g., centroids or medoids). Instead, the clustering results depend mutually on the sums of intracluster dissimilarity coefficients in all clusters. In other words, the objective function value to be minimized is the total intracluster dissimilarity.

As aforementioned in Section I, there is a shortcoming in the objective function in problem (7). As the total within-cluster dissimilarity is quadratically proportional to the cluster size, it would result in unnaturally balanced clusters. To illustrate this point, consider a simple case to cluster an even number of data into two clusters (i.e., n is even and $p = 2$). If the two clusters contain equal numbers of data

TABLE I

INFORMATION ON THE INTERNAL CLUSTER VALIDITY INDICES, WHERE ↑ AND ↓ INDICATE THEIR DESIRABILITY

Measure	acronym	definition	↑ or ↓	ref.
McClain-Rao Index	MRI	$N_B S_W / N_W S_B$	↓	[93]
G+ Index	GPI	$2s^- / (N_T(N_T - 1))$	↓	[94]
Baker-Hubert Gamma Index	BHGI	$(s^+ - s^-)/(s^+ + s^-)$	↑	[95]
C Index	CI	$(S_W - S_{\min})/(S_{\max} - S_{\min})$	↓	[96]
Tau Index	TI	$(s^+ - s^-)/(\sqrt{N_B N_W (N_T(N_T-1)/2)})$	↑	[97]
Dunn Generalized Index	DGI	$\min_{k \neq k'} \delta(C_k, C'_k) / \max_k \Delta(C_k)$	↑	[98]
Ratkowsky-Lance Index	RLI	$\sqrt{\sum_{j=1}^{m} (BGSS_j / TSS_j)/mp}$	↑	[99]
Calinski Harabasz Index	CHI	$(n-p)BGSS/(p-1)WGSS$	↑	[100]
Ray-Turi Index	RTI	$WGSS/(n \min_{k<k'} \|\|c_k - c_{k'}\|\|^2)$	↓	[101]
Wemmert-Gancarski Index	WGI	$\frac{1}{n}\sum_{k=1}^{p} \max\{0, n_k - \sum_{x \in C_k} (\|\|x - c_k\|\| / \min_{k \neq k'} \|\|x - c_{k'}\|\|)\}$	↑	[97]
Dunn Index	DI	$\min_{i,j}\{\min_{x \in C_i, y \in C_j} d(x,y)\} / \max_k \{\max_{x,y \in C_k} d(x,y)\}$	↑	[102]
Trace WiB Index	TWBI	$Tr(WG^{-1}.BG)$	↑	[103]
Ball-Hall Index	BHI	$\frac{1}{p}\sum_{k=1}^{p} \frac{1}{n_k} \sum_{x \in C_k} \|\|x - c_k\|\|^2$	↑	[104]
PBM Index	PBMI	$\left(\frac{1}{p} \times \frac{\sum_{x \in D} d(x,c)}{\sum_{k=1}^{p} \sum_{x \in C_k} d(x,c_k)} \times \max_{k<k'} d(c_k - c_{k'})\right)^2$	↑	[105]
Xie-Beni Index	XBI	$WGSS/[n \times \left(\min_{i<j}\{\min_{x \in C_i, y \in C_j} d(x,y)\}\right)^2]$	↓	[106]
Davies Bouldin Index	DBI	$\frac{1}{p}\sum_{i}^{n} \max_{j \neq i}\{\frac{1}{n_i}\sum_{x \in C_i} \|\|x - c_i\|\| + \frac{1}{n_j}\sum_{x \in C_j} \|\|x - c_j\|\| / d(c_i, c_j)\}$	↓	[107]
Det Ratio Index	DRI	$\det(T)/\det(WG)$	↓	[108]
Ksq DetW Index	KDWI	$p^2 \det(WG)$	↑	[109]
Log Det Ratio Index	LDRI	$n \log(\det(T)/\det(WG))$	↓	[108]
Log SS Ratio Index	LSSRI	$\log(BGSS/WGSS)$	↓	[110]
Trace W Index	TWI	$Tr(WG)$	↑	[111]

D: data set; c: center of D; C_i: the i-th cluster; n_i: the number of data in C_i; c_k: center of C_k; c: center of the whole set of data; $d(x,y)$:distance between x and y; S_W: the sum of the within-cluster distances $\sum_{k=1}^{p}\sum_{x,y \in C_k, x<y} d(x,y)$; S_B: the sum of the between-cluster distances $\sum_{k<k'}\sum_{x \in C_k, y \in C_{k'}, x<y} d(x,y)$; N_W: the total number of distances between pairs of points belonging to the same cluster; N_B: the total number of distances between pairs of points that do not belong to the same cluster; s^+: the number of times a distance between two points that belong to the same cluster is strictly smaller than the distance between two points not belonging to the same cluster; s^-: the number of times a distance between two points lying in the same cluster is strictly greater than a distance between two points not belonging to the same cluster; N_T: the total number of pairs of points in the dataset $n(n-1)/2$; $S_{\min}$: the smallest within-cluster distance; $S_{\max}$: the largest within-cluster distance; $\Delta(C_k)$: the within-cluster distances $\sum_{x,y \in C_k, x \neq y} d(x,y)/(n_k(n_k - 1))$; $\delta(C_k, C'_k)$: the between-cluster distance $\min_{x \in C_k, j \in C'_k} d(x,y)$; $BGSS_j$: $\sum_{k=1}^{p} n_k (c_k^j - c^j)^2$; TSS_j: $\sum_{i=1}^{N}(D_{ij} - c^j)$; $BGSS$: $\sum_{k=1}^{p} n_k \|\|c_k - c\|\|^2$; $WGSS$: $\sum_{k=1}^{p}\sum_{x \in C_k} \|\|x - c_k\|\|^2$; $T := t_{ij} = \sum_{x \in D}(x^i - c^i)(x^j - c^j)$; $WG^k \in \mathbb{R}^{p \times p}$ is the within-group scatter matrix of cluster k, and its elements $w_{ij}^k := n_k \mathrm{Cov}(x_k^i, x_k^j)$; $w_{ij}^k := n_k \mathrm{Var}(x_k^i)$, where x_k^i denotes the ith feature of a sample point in cluster k;$WG = \sum_{k=1}^{p} WG^k$;BG is between-group scatter matrix, and its element $b_{ij} = \sum_{k=1}^{p} n_k (c_k^i - c^i)(c_k^j - c^j)$.

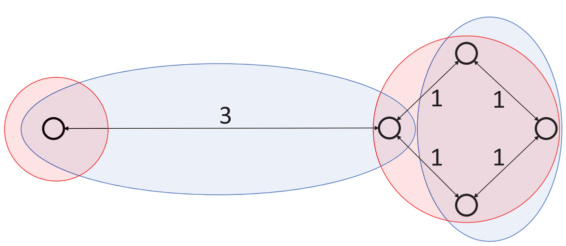

Fig. 1. Illustration of the clustering results based on the distance measures with and without normalization.

(i.e., $n_1 = n_2 = n/2$, where n_l denotes the number of data in cluster l), then the total number of weighted connections is up to $n^2/2$. Now suppose that cluster 1 contains q more data than cluster 2. Then, $n_1 = (n+q)/2$, $n_2 = (n-q)/2$, and the total number of weighted connections increases to $(n^2+q^2)/2$ (i.e., $q^2/2$ more). As a result, the clustering results based on the formulation in (7) tend to be unnaturally balanced in terms

TABLE II

PARAMETERS OF THE DATASETS (I.E., n, m, p, AND b), PROBLEM (15) (I.E., PENALTY PARAMETER ρ), BMM COOLING SCHEDULE (5) (I.E., T_0), AND CNO-CC (I.E., M AND N) USED IN THE EXPERIMENTS

Dataset	n	m	p	b	ρ	T_0	M	N
SJC1	100	2	10	720	5	2.6	500	8
SJC2	200	2	15	840	6	2.5	500	8
SJC3a	300	2	25	740	5	1.7	500	16
SJC3b	300	2	30	740	6.5	2.5	900	8
SJC4a	402	2	30	840	6.5	1.7	800	32
SJC4b	402	2	40	840	5.5	1.4	500	16
U724_010	724	2	10	4175	8	2.5	600	32
Doni1	1000	2	6	200	7	2.5	800	16
Rl1304_010	1304	2	10	7237	45	2.5	1000	64
Doni2	2000	2	6	400	7	2.5	1000	64

TABLE III

MEAN VALUES AND STANDARD DEVIATIONS OF 21 INTERNAL CLUSTER VALIDITY CRITERIA RESULTING FROM CNO-CC AND SEVEN BASELINES ON SJC1 ($n = 100,\ p = 10$, AND $b = 720$), SJC2 ($n = 200,\ p = 15$, AND $b = 840$), SJC3a ($n = 300,\ p = 25$, AND $b = 740$), AND SJC3b ($n = 300,\ p = 20$, AND $b = 740$)

SJC1	CNO-CC	TS [20]	GRASP-VND [20]	GRASP-VND-TS [20]	IVNS [21]	GVNS [25]	SGVNS [25]	NDVNS [26]
MRI↓	0.2912 ± 0.0061	0.3185 ± 0.0126	0.3031 ± 0.0101	0.3043 ± 0.0060	0.2898 ± 0.0055	**0.2885 ± 0.0057**	0.2907 ± 0.0061	0.2998 ± 0.0082
GPI↓	0.0089 ± 0.0007	0.0120 ± 0.0015	0.0103 ± 0.0011	0.0104 ± 0.0006	0.0087 ± 0.0006	**0.0086 ± 0.0006**	0.0088 ± 0.0007	0.0098 ± 0.0009
BHGI↑	**0.8999 ± 0.0080**	0.8606 ± 0.0167	0.8784 ± 0.0136	0.8775 ± 0.0080	0.8965 ± 0.0080	0.8985 ± 0.0083	0.8954 ± 0.0087	0.8841 ± 0.0110
CI↓	**0.0448 ± 0.0033**	0.0613 ± 0.0069	0.0535 ± 0.0056	0.0540 ± 0.0034	0.0461 ± 0.0033	0.0453 ± 0.0034	0.0466 ± 0.0036	0.0514 ± 0.0045
TI↑	**0.3790 ± 0.0049**	0.3564 ± 0.0069	0.3608 ± 0.0059	0.3608 ± 0.0044	0.3686 ± 0.0042	0.3696 ± 0.0045	0.3680 ± 0.0047	0.3644 ± 0.0054
DGI↑	0.3736±0.0410	0.2722 ± 0.0744	0.3519 ± 0.0543	0.3293 ± 0.0923	**0.3925 ± 0.0328**	0.3819 ± 0.0353	0.3766 ± 0.0302	0.3015 ± 0.0793
RLI↑	**0.3019 ± 0.0008**	0.2970 ± 0.0034	0.2998 ± 0.0015	0.2990 ± 0.0024	0.3017 ± 0.0006	0.3018 ± 0.0006	0.3016 ± 0.0006	0.2994 ± 0.0024
CHI↑	**102.4833 ± 5.3809**	76.3844 ± 12.2112	88.1989 ± 8.4837	84.1652 ± 9.9181	99.7049 ± 4.7139	100.8864 ± 4.8262	99.2471 ± 4.7849	86.8653 ± 11.7945
RTI↓	**0.3876 ± 0.0834**	1.2469 ± 1.4654	55.6331 ± 242.9702	0.8844 ± 0.9473	0.4567 ± 0.3158	0.3889 ± 0.2383	0.4203 ± 0.2383	1.4253 ± 3.1528
WGI↑	0.4819 ± 0.0202	0.388 ± 0.0331	0.4398 ± 0.0355	0.4311 ± 0.0305	0.4844 ± 0.0175	**0.4878 ± 0.0185**	0.4819 ± 0.0202	0.4468 ± 0.0310
DI↑	**0.0799 ± 0.0095**	0.0582 ± 0.0151	0.0739 ± 0.0099	0.0676 ± 0.0180	0.0789 ± 0.0073	0.0772 ± 0.0074	0.0747 ± 0.0071	0.0630 ± 0.0149
TWBI↑	19.9010 ± 1.3024	14.8238 ± 2.3890	18.0611 ± 1.6630	16.8894 ± 2.2708	19.9329 ± 0.4225	**20.015 ± 0.3696**	19.5196 ± 0.8641	17.154 ± 2.1965
BHI↑	0.0089 ± 0.0006	0.0131 ± 0.0036	0.0107 ± 0.0012	**0.0118 ± 0.0031**	0.0094 ± 0.0004	0.0094 ± 0.0004	0.0095 ± 0.0004	0.0117 ± 0.0031
PBMI↑	0.1055 ± 0.0054	0.0792 ± 0.0113	0.0953 ± 0.0096	0.0910 ± 0.0099	0.1072 ± 0.0043	**0.1083 ± 0.0042**	0.1065 ± 0.0049	0.0950 ± 0.0094
XBI↓	**9.0244 ± 1.4727**	14.1175 ± 5.3499	10.1555 ± 1.9040	10.7707 ± 2.9138	9.0814 ± 1.5593	9.5577 ± 1.5712	9.9171 ± 1.3681	11.0226 ± 2.1915
DBI↓	**0.8325 ± 0.0442**	1.2400 ± 0.4275	1.8042 ± 3.3009	1.0808 ± 0.2865	0.8628 ± 0.0878	0.8417 ± 0.0729	0.8573 ± 0.0715	1.1279 ± 0.3939
DRI↑	**98.7081 ± 8.8463**	60.6881 ± 14.4335	78.4804 ± 13.0010	72.4738 ± 12.2052	94.7907 ± 8.8444	97.2317 ± 8.8509	94.7061 ± 8.6671	77.3635 ± 15.9286
KDWI↓	**18.2605 ± 1.6466**	31.5014 ± 9.4670	23.4615 ± 4.3656	25.5810 ± 5.7475	19.0216 ± 1.7137	18.5441 ± 1.7129	19.0367 ± 1.7296	24.2377 ± 6.0042
LDRI↑	**458.8339 ± 8.9885**	407.5064 ± 26.4344	434.8922 ± 17.4596	426.7030 ± 19.3799	454.7617 ± 9.1900	457.3108 ± 9.1915	454.6813 ± 9.1303	432.6364 ± 22.2501
LSSRI↑	**2.3258 ± 0.0535**	2.0195 ± 0.1753	2.1725 ± 0.0983	2.1226 ± 0.1317	2.2986 ± 0.0472	2.3103 ± 0.0482	2.2939 ± 0.0480	2.1523 ± 0.1443
TWI↓	**0.8889 ± 0.0440**	1.1808 ± 0.1958	1.0234 ± 0.0918	1.0731 ± 0.1376	0.9110 ± 0.0390	0.9013 ± 0.0398	0.9148 ± 0.0397	1.0462 ± 0.1425

SJC2	CNO-CC	TS [20]	GRASP-VND [20]	GRASP-VND-TS [20]	IVNS [21]	GVNS [25]	SGVNS [25]	NDVNS [26]
MRI↓	**0.2196 ± 0.0008**	0.2368 ± 0.0046	0.2294 ± 0.0042	0.2275 ± 0.0041	0.2231 ± 0.0021	0.2237 ± 0.0015	0.2214 ± 0.0009	0.2260 ± 0.0025
GPI↓	**0.0030 ± 0.0001**	0.0042 ± 0.0003	0.0037 ± 0.0003	0.0036 ± 0.0003	0.0033 ± 0.0002	0.0033 ± 0.0001	0.0031 ± 0.0001	0.0035 ± 0.0002
BHGI↑	**0.9505 ± 0.0012**	0.9302 ± 0.0056	0.9375 ± 0.0048	0.9398 ± 0.0045	0.9443 ± 0.0027	0.9440 ± 0.0018	0.9465 ± 0.0011	0.9411 ± 0.0030
CI↓	**0.0290 ± 0.0006**	0.0398 ± 0.0027	0.0358 ± 0.0023	0.0347 ± 0.0022	0.0324 ± 0.0012	0.0326 ± 0.0009	0.0314 ± 0.0006	0.0340 ± 0.0014
TI↑	**0.3313 ± 0.0021**	0.3208 ± 0.0025	0.3220 ± 0.0017	0.3231 ± 0.0017	0.3238 ± 0.0008	0.3243 ± 0.0011	0.3246 ± 0.0006	0.3229 ± 0.0013
DGI↑	**0.4445 ± 0.0363**	0.2543 ± 0.1119	0.3048 ± 0.1258	0.3231 ± 0.0822	0.3590 ± 0.0697	0.3875 ± 0.0681	0.3664 ± 0.0507	0.3553 ± 0.0788
RLI↑	**0.2513 ± 0.0002**	0.2492 ± 0.0010	0.2497 ± 0.0011	0.2502 ± 0.0008	0.2506 ± 0.0006	0.2507 ± 0.0002	0.2510 ± 0.0002	0.2503 ± 0.0006
CHI↑	**258.5375 ± 3.3617**	198.1424 ± 25.4384	210.2064 ± 30.5971	223.6604 ± 23.6017	237.8632 ± 20.5717	241.5310 ± 6.6453	249.4157 ± 3.0561	229.6104 ± 19.1541
RTI↓	**0.2915 ± 0.0265**	1.4581 ± 3.6889	0.8145 ± 1.3649	0.5058 ± 0.4307	0.3317 ± 0.0644	0.3641 ± 0.0772	0.3172 ± 0.0221	0.5341 ± 0.6963
WGI↑	**0.5047 ± 0.0140**	0.4345 ± 0.0206	0.4619 ± 0.0211	0.4713 ± 0.0191	0.4850 ± 0.0123	0.4836 ± 0.0115	0.4899 ± 0.0093	0.4761 ± 0.0130
DI↑	**0.1036 ± 0.0068**	0.0543 ± 0.0257	0.0605 ± 0.0298	0.0666 ± 0.0203	0.0747 ± 0.0186	0.0829 ± 0.0154	0.0852 ± 0.0116	0.0702 ± 0.0211
TWBI↑	**39.8867 ± 1.3068**	30.7008 ± 3.9930	32.8270 ± 4.7591	34.9680 ± 4.3016	37.3805 ± 3.3259	37.8883 ± 2.1739	38.8283 ± 0.8177	36.4887 ± 3.3470
BHI↑	0.0065 ± 0.0001	**0.0092 ± 0.0021**	0.0088 ± 0.0020	0.0081 ± 0.0017	0.0073 ± 0.0010	0.0071 ± 0.0003	0.0069 ± 0.0001	0.0077 ± 0.0013
PBMI↑	**0.0882 ± 0.0033**	0.0737 ± 0.0051	0.0758 ± 0.0052	0.0798 ± 0.0046	0.0812 ± 0.0036	0.0829 ± 0.0049	0.0818 ± 0.0010	0.0800 ± 0.0040
XBI↓	**7.5351 ± 1.2285**	98.7739 ± 258.4076	88.2315 ± 237.7553	11.1735 ± 2.8489	10.9023 ± 2.9168	10.0255 ± 2.9901	11.4161 ± 2.5194	10.4440 ± 2.8216
DBI↓	**0.8143 ± 0.0189**	1.0817 ± 0.4001	1.0054 ± 0.2123	0.9441 ± 0.1134	0.8619 ± 0.0455	0.8764 ± 0.0384	0.8448 ± 0.0105	0.9206 ± 0.1464
DRI↑	**394.4584 ± 10.9094**	244.0968 ± 52.0090	277.7956 ± 65.5524	307.0154 ± 56.5147	343.1438 ± 44.8690	350.5524 ± 20.1827	369.8853 ± 8.269	322.0058 ± 44.7075
KDWI↓	**94.7196 ± 2.6533**	160.8219 ± 39.6709	143.0536 ± 39.0039	126.6106 ± 29.1817	111.3756 ± 20.9530	106.8548 ± 6.4191	100.9878 ± 2.3516	118.9380 ± 23.0074
LDRI↑	**1195.4295 ± 5.5669**	1094.7444 ± 46.0297	1119.4389 ± 51.4379	1141.6510 ± 41.1824	1165.5769 ± 31.2172	1171.5785 ± 11.7554	1182.5898 ± 4.5613	1152.6488 ± 32.6283
LSSRI↑	**2.9737 ± 0.0131**	2.6992 ± 0.1360	2.7560 ± 0.1537	2.8231 ± 0.1123	2.8862 ± 0.0981	2.9053 ± 0.0280	2.9377 ± 0.0124	2.8514 ± 0.0913
TWI↓	**1.3040 ± 0.0163**	1.7021 ± 0.2272	1.6173 ± 0.2428	1.5117 ± 0.1692	1.4226 ± 0.1482	1.3917 ± 0.0375	1.3492 ± 0.0162	1.4694 ± 0.1377

SJC3a	CNO-CC	TS [20]	GRASP-VND [20]	GRASP-VND-TS [20]	IVNS [21]	GVNS [25]	SGVNS [25]	NDVNS [26]
MRI↓	0.1708 ± 0.0013	0.1841 ± 0.0030	0.1753 ± 0.0025	0.1756 ± 0.0023	**0.1700 ± 0.0010**	0.1723 ± 0.0015	0.1712 ± 0.0009	0.1739 ± 0.0018
GPI↓	**0.0012 ± 0.0000**	0.0017 ± 0.0001	0.0014 ± 0.0001	0.0014 ± 0.0001	**0.0012 ± 0.0000**	0.0013 ± 0.0001	**0.0012 ± 0.0000**	0.0013 ± 0.0001
BHGI↑	**0.9680 ± 0.0012**	0.954 ± 0.0032	0.962 ± 0.0024	0.9616 ± 0.0025	0.9669 ± 0.0012	0.9649 ± 0.0017	0.9659 ± 0.001	0.9632 ± 0.0018
CI↓	**0.0210 ± 0.0007**	0.0289 ± 0.0015	0.0244 ± 0.0013	0.0246 ± 0.0012	0.0218 ± 0.0005	0.0230 ± 0.0008	0.0224 ± 0.0005	0.0238 ± 0.0009
TI↑	**0.2653 ± 0.0019**	0.2572 ± 0.0012	0.259 ± 0.0008	0.2588 ± 0.0011	0.2597 ± 0.0005	0.2592 ± 0.0008	0.2594 ± 0.0007	0.2588 ± 0.0007
DGI↑	**0.3717 ± 0.0504**	0.1955 ± 0.0875	0.2586 ± 0.0889	0.2742 ± 0.0692	0.3501 ± 0.0704	0.3079 ± 0.049	0.3390 ± 0.0439	0.2854 ± 0.0543
RLI↑	**0.1970 ± 0.0001**	0.196 ± 0.0006	0.1965 ± 0.0003	0.1965 ± 0.0004	0.1969 ± 0.0003	0.1968 ± 0.0002	0.1969 ± 0.0001	0.1966 ± 0.0002
CHI↑	**404.7875 ± 6.9407**	302.7338 ± 35.5859	344.7671 ± 33.9902	347.9867 ± 32.7213	387.5973 ± 25.0785	379.5439 ± 15.2302	387.4522 ± 7.5671	364.6832 ± 16.9603
RTI↓	**0.3630 ± 0.1204**	1.1273 ± 1.6790	1.2072 ± 2.678	0.5998 ± 0.3756	0.6124 ± 0.6477	0.4331 ± 0.1142	0.4929 ± 0.2348	0.6200 ± 0.5232
WGI↑	**0.5066 ± 0.0060**	0.4231 ± 0.0126	0.467 ± 0.0144	0.4666 ± 0.0120	0.4911 ± 0.0099	0.4825 ± 0.0072	0.4873 ± 0.0046	0.4751 ± 0.0115
DI↑	**0.0915 ± 0.0114**	0.0418 ± 0.0187	0.053 ± 0.0183	0.0526 ± 0.0154	0.0731 ± 0.0172	0.0637 ± 0.0175	0.0704 ± 0.0141	0.0577 ± 0.0177
TWBI↑	**72.0777 ± 2.2505**	54.9746 ± 6.9229	62.0912 ± 6.9259	62.0780 ± 5.2805	69.5166 ± 3.4699	68.0131 ± 2.5883	69.3132 ± 2.1077	66.1876 ± 2.8711
BHI↑	0.0036 ± 0.0001	**0.0055 ± 0.0015**	0.0048 ± 0.0011	0.0045 ± 0.0009	0.0040 ± 0.0006	0.0040 ± 0.0002	0.0039 ± 0.0001	0.0042 ± 0.0003
PBMI↑	**0.0634 ± 0.0018**	0.0510 ± 0.0029	0.0571 ± 0.004	0.0583 ± 0.0027	0.0624 ± 0.0022	0.0605 ± 0.0018	0.0618 ± 0.0012	0.0593 ± 0.0020
XBI↓	9.3436 ± 2.0224	18.3319 ± 7.0728	10.4922 ± 1.9231	11.3083 ± 2.2616	**8.0296 ± 2.3749**	11.3727 ± 2.5176	9.3656 ± 2.4405	10.9898 ± 2.6433
DBI↓	**0.8200 ± 0.0216**	1.1033 ± 0.2320	1.0189 ± 0.259	0.9311 ± 0.0661	0.8987 ± 0.1277	0.8755 ± 0.0355	0.8684 ± 0.0278	0.9115 ± 0.0806
DRI↑	**1100.1036 ± 36.3253**	646.8354 ± 133.3649	827.5845 ± 131.2393	835.3834 ± 131.4999	1021.8948 ± 109.4389	976.6157 ± 70.6804	1012.7544 ± 37.0671	904.9806 ± 77.1335
KDWI↓	**182.2096 ± 6.0241**	325.4795 ± 84.3851	248.8755 ± 46.6535	246.5349 ± 46.3217	198.8599 ± 28.8223	206.1698 ± 16.5549	197.9757 ± 7.3615	222.9684 ± 21.1989
LDRI↑	**2100.7924 ± 9.9122**	1934.6818 ± 68.8172	2011.5618 ± 51.6638	2014.3997 ± 51.6033	2076.8366 ± 37.2388	2064.4296 ± 22.8465	2075.9358 ± 11.0679	2041.2665 ± 26.9296
LSSRI↑	**3.5645 ± 0.0172**	3.2671 ± 0.1241	3.3991 ± 0.1049	3.4089 ± 0.0995	3.5190 ± 0.0707	3.4995 ± 0.0412	3.5207 ± 0.0196	3.4592 ± 0.0478
TWI↓	**1.0829 ± 0.0181**	1.4538 ± 0.1831	1.2774 ± 0.1374	1.2647 ± 0.1282	1.1343 ± 0.0848	1.1542 ± 0.0474	1.1300 ± 0.0216	1.2005 ± 0.0572

SJC3b	CNO-CC	TS [20]	GRASP-VND [20]	GRASP-VND-TS [20]	IVNS [21]	GVNS [25]	SGVNS [25]	NDVNS [26]
MRI↓	0.1944 ± 0.0019	0.2084 ± 0.0031	0.2025 ± 0.0050	0.2031 ± 0.0046	**0.1936 ± 0.0022**	0.1962 ± 0.0022	0.1937 ± 0.0032	0.1989 ± 0.0022
GPI↓	**0.0020 ± 0.0001**	0.0027 ± 0.0002	0.0025 ± 0.0002	0.0025 ± 0.0002	**0.0020 ± 0.0001**	**0.0022 ± 0.0001**	**0.002 ± 0.0001**	0.0023 ± 0.0001
BHGI↑	**0.9567 ± 0.0020**	0.9407 ± 0.0034	0.9463 ± 0.0050	0.9456 ± 0.0050	0.9555 ± 0.0022	0.9528 ± 0.0023	0.9553 ± 0.0029	0.9503 ± 0.0022
CI↓	**0.0263 ± 0.0011**	0.0347 ± 0.0017	0.0317 ± 0.0026	0.0320 ± 0.0024	0.0270 ± 0.0011	0.0284 ± 0.0012	0.0270 ± 0.0016	0.0297 ± 0.0011
TI↑	**0.2939 ± 0.0024**	0.2858 ± 0.0017	0.2864 ± 0.0013	0.2863 ± 0.0017	0.2888 ± 0.0007	0.2880 ± 0.0009	0.2888 ± 0.0007	0.2878 ± 0.0010
DGI↑	**0.3511 ± 0.0375**	0.1351 ± 0.0747	0.2280 ± 0.0912	0.2047 ± 0.0810	0.2570 ± 0.0736	0.2478 ± 0.0810	0.2434 ± 0.0532	0.1990 ± 0.0883
RLI↑	**0.2190 ± 0.0001**	0.2172 ± 0.0006	0.2180 ± 0.0007	0.2178 ± 0.0007	0.2188 ± 0.0004	0.2185 ± 0.0005	0.2186 ± 0.0006	0.2182 ± 0.0006
CHI↑	**371.6300 ± 9.3640**	266.1958 ± 28.3107	306.6767 ± 37.512	301.2459 ± 38.7819	354.9103 ± 30.2928	336.7602 ± 27.6364	349.9783 ± 30.5821	320.2670 ± 33.5022
RTI↓	**0.3934 ± 0.0867**	26.8980 ± 106.1727	0.9980 ± 1.0222	6.6252 ± 24.4985	0.6246 ± 1.0461	0.5880 ± 0.3407	0.5670 ± 0.5334	0.5195 ± 0.1956
WGI↑	0.4739 ± 0.0144	0.4069 ± 0.0180	0.4208 ± 0.0278	0.4247 ± 0.0218	**0.4763 ± 0.0097**	0.4613 ± 0.0141	0.4735 ± 0.0156	0.4466 ± 0.0134
DI↑	**0.0705 ± 0.0138**	0.0252 ± 0.0133	0.0420 ± 0.0179	0.0348 ± 0.0149	0.0447 ± 0.0144	0.0451 ± 0.0188	0.0427 ± 0.0150	0.0382 ± 0.0176
TWBI↑	**50.9202 ± 1.8003**	38.0629 ± 5.3262	42.9237 ± 5.9219	42.4517 ± 5.2993	50.2138 ± 3.5694	48.3719 ± 3.3310	50.1709 ± 2.1726	45.0213 ± 4.4839
BHI↑	0.0050 ± 0.0002	0.0084 ± 0.0020	**0.0068 ± 0.0016**	**0.0068 ± 0.0014**	0.0056 ± 0.0010	0.0060 ± 0.0009	0.0057 ± 0.0009	0.0066 ± 0.0015
PBMI↑	**0.0686 ± 0.0021**	0.0562 ± 0.0038	0.0612 ± 0.0051	0.0614 ± 0.0050	0.0683 ± 0.0029	0.0663 ± 0.0027	0.0684 ± 0.0030	0.0638 ± 0.0033
XBI↓	**11.7227 ± 2.8035**	166.3153 ± 343.7722	∞ ± NaN	121.7859 ± 316.5259	54.2989 ± 176.7933	56.1617 ± 185.2908	16.1663 ± 3.1423	102.1993 ± 255.3857
DBI↓	**0.8754 ± 0.0291**	1.7279 ± 1.9800	1.0756 ± 0.1909	1.2098 ± 0.7705	0.9351 ± 0.1640	0.9458 ± 0.0669	0.9305 ± 0.0895	1.0084 ± 0.1657
DRI↑	**574.9973 ± 29.9329**	316.2264 ± 59.0211	410.0772 ± 83.8439	395.9581 ± 85.6028	535.1819 ± 70.7836	486.9883 ± 65.3027	522.5146 ± 71.7992	440.3768 ± 77.6892
KDWI↓	**223.4398 ± 11.3645**	419.3810 ± 81.5267	326.8737 ± 76.2076	341.4278 ± 89.9719	244.5495 ± 40.8939	268.2979 ± 40.8280	250.1556 ± 37.5735	301.1391 ± 60.7441
LDRI↑	**1905.9296 ± 15.4340**	1721.8702 ± 57.0050	1798.4602 ± 65.2702	1786.9037 ± 71.1263	1881.8742 ± 44.5762	1853.6927 ± 42.8079	1874.7225 ± 43.2939	1821.3874 ± 57.0240
LSSRI↑	**3.2272 ± 0.0251**	2.8885 ± 0.1068	3.0279 ± 0.1279	3.0091 ± 0.1367	3.1777 ± 0.0921	3.1257 ± 0.0852	3.1637 ± 0.0901	3.0733 ± 0.1088
TWI↓	**1.5009 ± 0.0360**	2.0831 ± 0.2115	1.8287 ± 0.2323	1.8638 ± 0.2572	1.5795 ± 0.1498	1.6593 ± 0.1401	1.6004 ± 0.1419	1.7478 ± 0.1873

of resulting cluster sizes. Consider a simple example illustrated in Fig. 1, where five samples are to be partitioned into two clusters. If the total within-cluster Euclidean distance is used as the objective function, then the total within-cluster distance of the resulting clusters surrounded by the two ellipses is $5+\sqrt{2}$ (≈6.414), whereas more natural clusters surrounded by the two circles result in a bigger total within-cluster distance (i.e., $4 + 2\sqrt{2} \approx 6.828$).

The unnatural clustering result is because the sum of dissimilarity coefficients in each cluster is not normalized by the size of the cluster, and the terms in the objective function (i.e., the numbers of intra-cluster weighted con-

TABLE IV

MEAN VALUES AND STANDARD DEVIATIONS OF 21 INTERNAL CLUSTER VALIDITY CRITERIA RESULTING FROM CNO-CC AND SEVEN BASELINES ON SJC4a ($n = 402,\ p = 30$, AND $b = 840$), SJC4b ($n = 402,\ p = 40$, AND $b = 840$), DONI1 ($n = 1000,\ p = 6$, AND $b = 200$), AND U724_010 ($n = 724,\ p = 10$, AND $b = 4175$)

SJC4a	CNO-CC	TS [20]	GRASP-VND [20]	GRASP-VND-TS [20]	IVNS [21]	GVNS [25]	SGVNS [25]	NDVNS [26]
MRI↓	0.1617 ± 0.0009	0.1745 ± 0.0034	0.1656 ± 0.0024	0.1662 ± 0.0026	**0.1608 ± 0.0006**	0.1630 ± 0.0016	0.1609 ± 0.0011	0.1652 ± 0.0023
GPI↓	**0.0009 ± 0.0000**	0.0012 ± 0.0001	0.0010 ± 0.0001	0.0010 ± 0.0001	**0.0009 ± 0.0000**	0.0009 ± 0.0001	**0.0009 ± 0.0000**	0.0010 ± 0.0001
BHGI↑	**0.9722 ± 0.0007**	0.9597 ± 0.0034	0.9675 ± 0.0024	0.9669 ± 0.0030	0.9716 ± 0.0007	0.9700 ± 0.0017	0.9717 ± 0.0009	0.9676 ± 0.0024
CI↓	**0.0203 ± 0.0005**	0.0278 ± 0.0017	0.0232 ± 0.0013	0.0235 ± 0.0014	0.0208 ± 0.0004	0.0218 ± 0.0009	0.0208 ± 0.0006	0.0230 ± 0.0013
TI↑	**0.2445 ± 0.0018**	0.2380 ± 0.0010	0.2399 ± 0.0013	0.2396 ± 0.0012	0.2401 ± 0.0006	0.2404 ± 0.0009	0.2405 ± 0.0004	0.2395 ± 0.0011
DGI↑	**0.3219 ± 0.0371**	0.1878 ± 0.0818	0.2701 ± 0.0642	0.2859 ± 0.0444	0.3195 ± 0.0166	0.3090 ± 0.0546	0.3135 ± 0.0243	0.2738 ± 0.063
RLI↑	**0.1802 ± 0.0000**	0.1793 ± 0.0005	0.1799 ± 0.0002	0.1798 ± 0.0003	**0.1802 ± 0.0000**	0.1801 ± 0.0002	**0.1802 ± 0.0000**	0.1799 ± 0.0002
CHI↑	**498.6002 ± 6.0055**	369.2308 ± 52.5659	443.6728 ± 32.5099	441.9115 ± 37.6014	493.8303 ± 7.7804	474.4083 ± 31.0926	494.3243 ± 8.7970	444.7208 ± 38.7877
RTI↓	0.3992 ± 0.0771	1.2593 ± 2.3404	0.6141 ± 0.4526	0.5599 ± 0.2479	0.3759 ± 0.0618	0.4317 ± 0.0818	**0.3529 ± 0.0752**	0.4639 ± 0.1143
WGI↑	**0.4956 ± 0.0072**	0.4146 ± 0.0179	0.4571 ± 0.0144	0.4552 ± 0.0138	0.4884 ± 0.0038	0.4794 ± 0.0091	0.4892 ± 0.0099	0.4636 ± 0.0157
DI↑	**0.0706 ± 0.012**	0.0361 ± 0.0169	0.0502 ± 0.0138	0.0528 ± 0.0152	0.0628 ± 0.0083	0.0639 ± 0.0145	0.0639 ± 0.0065	0.0481 ± 0.0145
TWBI↑	**78.5796 ± 1.0125**	59.2018 ± 7.5660	70.0792 ± 4.8002	70.3298 ± 4.3991	77.9525 ± 1.3813	74.732 ± 5.0563	77.6053 ± 1.5804	70.5927 ± 5.5945
BHI↑	0.0033 ± 0.0001	**0.0052 ± 0.0015**	0.0040 ± 0.0006	0.0039 ± 0.0004	0.0034 ± 0.0001	0.0036 ± 0.0005	0.0034 ± 0.0001	0.0040 ± 0.0007
PBMI↑	0.0463 ± 0.0011	0.0378 ± 0.0031	0.0432 ± 0.0019	0.0434 ± 0.0021	0.0468 ± 0.0011	0.0457 ± 0.0012	**0.0470 ± 0.0011**	0.0439 ± 0.0019
XBI↓	12.4057 ± 2.6121	56.6422 ± 166.5107	13.8765 ± 3.2790	13.7061 ± 3.0683	11.6069 ± 0.7190	12.5355 ± 2.8518	**12.3814 ± 2.3056**	13.5358 ± 2.2493
DBI↓	0.8499 ± 0.0235	1.1145 ± 0.4061	0.9412 ± 0.0792	0.9223 ± 0.0534	0.8524 ± 0.0157	0.8764 ± 0.0430	**0.8393 ± 0.0238**	0.9178 ± 0.0667
DRI↑	1544.5359 ± 35.9685	893.5758 ± 218.3921	1245.1532 ± 165.9415	1237.3347 ± 178.8119	1517.4247 ± 46.4320	1410.7967 ± 160.0167	1518.5656 ± 50.9333	**1252.9298 ± 185.1040**
KDWI↓	**405.7688 ± 9.3628**	747.2372 ± 204.5497	513.1724 ± 81.2655	518.8592 ± 93.3179	413.1743 ± 12.6167	452.4415 ± 79.659	412.9747 ± 15.0045	513.2491 ± 96.7607
LDRI↑	**2951.5735 ± 9.3185**	2719.3315 ± 104.4913	2861.2945 ± 58.0770	2857.9441 ± 64.5988	2944.379 ± 12.2879	2912.0389 ± 56.1234	2944.6335 ± 14.0181	2862.7154 ± 66.6842
LSSRI↑	**3.6601 ± 0.012**	3.3497 ± 0.1480	3.5408 ± 0.0766	3.5358 ± 0.0908	3.6505 ± 0.0157	3.6081 ± 0.0733	3.6514 ± 0.0181	3.5419 ± 0.0933
TWI↓	**1.3457 ± 0.0157**	1.8363 ± 0.2714	1.5154 ± 0.1179	1.5245 ± 0.1431	1.3585 ± 0.0208	1.4192 ± 0.1133	1.3573 ± 0.0244	1.5157 ± 0.1467
SJC4b	CNO-CC	TS [20]	GRASP-VND [20]	GRASP-VND-TS [20]	IVNS [21]	GVNS [25]	SGVNS [25]	NDVNS [26]
MRI↓	**0.1404 ± 0.0007**	0.1525 ± 0.0019	0.1421 ± 0.0010	0.1420 ± 0.0012	0.1387 ± 0.0004	0.1409 ± 0.0011	0.1412 ± 0.0015	0.1433 ± 0.0019
GPI↓	**0.0005 ± 0.0000**	0.0007 ± 0.0000	**0.0005 ± 0.0000**	**0.0005 ± 0.0000**	**5e-04 ± 0.0000**	**0.0005 ± 0.0000**	**0.0005 ± 0.0000**	**0.0005 ± 0.0000**
BHGI↑	**0.9789 ± 0.0005**	0.9687 ± 0.0018	0.9766 ± 0.0009	0.9767 ± 0.0009	**0.9789 ± 0.0004**	0.9776 ± 0.0007	0.9774 ± 0.0011	0.9760 ± 0.0016
CI↓	**0.0169 ± 0.0003**	0.0239 ± 0.0010	0.0186 ± 0.0005	0.0186 ± 0.0006	0.0170 ± 0.0003	0.0180 ± 0.0005	0.0181 ± 0.0008	0.0188 ± 0.0010
TI↑	**0.2119 ± 0.0016**	0.2058 ± 0.0009	0.2073 ± 0.0010	0.2072 ± 0.0007	0.2072 ± 0.0004	0.2074 ± 0.0007	0.2075 ± 0.0007	0.2091 ± 0.0011
DGI↑	0.3799 ± 0.0312	0.1994 ± 0.1107	0.3497 ± 0.0385	0.3406 ± 0.0548	0.3363 ± 0.0349	0.3514 ± 0.0396	0.3148 ± 0.1175	**0.3516 ± 0.0507**
RLI↑	**0.1566 ± 0.0000**	0.1562 ± 0.0002	0.1565 ± 0.0000	**0.1566 ± 0.0000**	0.1566 ± 0.0000	**0.1566 ± 0.0000**	**0.1566 ± 0.0000**	0.1565 ± 0.0001
CHI↑	**513.725 ± 5.4451**	384.1615 ± 47.2820	480.8249 ± 10.8762	484.1197 ± 11.2747	512.7677 ± 6.3912	495.8291 ± 8.0646	492.8008 ± 13.3102	475.9606 ± 19.9467
RTI↓	0.3774 ± 0.0441	0.8497 ± 0.9694	0.4767 ± 0.1459	0.4276 ± 0.0645	**0.3184 ± 0.0202**	0.3836 ± 0.0707	0.3860 ± 0.0591	0.4761 ± 0.1414
WGI↑	**0.5063 ± 0.0056**	0.4191 ± 0.0112	0.4785 ± 0.0104	0.4814 ± 0.0086	0.4991 ± 0.0025	0.4877 ± 0.0086	0.4824 ± 0.0111	0.4825 ± 0.0113
DI↑	**0.0935 ± 0.0116**	0.0435 ± 0.0248	0.0787 ± 0.0120	0.0805 ± 0.0156	0.0853 ± 0.0098	0.0872 ± 0.0114	0.0764 ± 0.0295	0.0807 ± 0.0170
TWBI↑	111.8347 ± 1.7476	83.8534 ± 10.3860	105.5817 ± 3.2580	105.7676 ± 2.7138	**113.6038 ± 1.6639**	108.7719 ± 3.0278	108.1956 ± 3.2120	104.3558 ± 4.4165
BHI↑	0.0023 ± 0.0000	**0.0036 ± 0.0010**	0.0026 ± 0.0001	0.0026 ± 0.0004	0.0024 ± 0.0000	0.0025 ± 0.0000	0.0025 ± 0.0001	0.0026 ± 0.0001
PBMI↑	**0.0388 ± 0.0014**	0.0308 ± 0.0017	0.0358 ± 0.0008	0.0359 ± 0.0008	0.0372 ± 0.0006	0.0366 ± 0.0010	0.0362 ± 0.0012	0.0355 ± 0.0011
XBI↓	**7.7450 ± 0.8843**	∞ ± NaN	8.9076 ± 2.0437	9.6219 ± 2.5903	9.4975 ± 1.5049	8.9239 ± 1.7979	∞ ± NaN	8.6996 ± 2.6812
DBI↓	**0.8171 ± 0.0154**	1.0310 ± 0.1911	0.8820 ± 0.0265	0.8704 ± 0.0208	0.8319 ± 0.0146	0.8514 ± 0.0221	0.8579 ± 0.0147	0.8797 ± 0.0313
DRI↑	3085.3363 ± 64.6932	1787.1545 ± 378.1010	2721.2710 ± 125.5754	2750.2717 ± 124.7211	**3095.0143 ± 74.1641**	2891.1427 ± 103.2887	2853.8801 ± 151.9160	2669.9930 ± 211.5043
KDWI↓	361.0846 ± 7.4726	659.5691 ± 184.6830	410.0756 ± 19.4624	405.7197 ± 18.8605	**360.0034 ± 8.6210**	385.6469 ± 13.7859	391.2915 ± 21.4836	419.6211 ± 33.8974
LDRI↑	3229.7520 ± 8.3742	2999.9735 ± 98.3293	3178.9453 ± 18.8101	3183.2219 ± 18.4518	**3230.9846 ± 9.6302**	3203.4579 ± 14.3661	3197.9354 ± 21.7116	3170.5005 ± 32.1289
LSSRI↑	**4.0136 ± 0.0106**	3.7150 ± 0.1332	3.9472 ± 0.0228	3.9540 ± 0.0234	4.0117 ± 0.0125	3.9780 ± 0.0163	3.9717 ± 0.0272	3.9364 ± 0.0421
TWI↓	**0.9522 ± 0.0098**	1.2860 ± 0.1796	1.0165 ± 0.0228	1.0097 ± 0.0233	0.9540 ± 0.0117	0.9860 ± 0.0158	0.9924 ± 0.0268	1.0278 ± 0.0427
Doni1	CNO-CC	TS [20]	GRASP-VND [20]	GRASP-VND-TS [20]	IVNS [21]	GVNS [25]	SGVNS [25]	NDVNS [26]
MRI↓	0.3972 ± 0.0030	0.3701 ± 0.0064	0.3703 ± 0.0069	0.368 ± 0.0057	0.3657 ± 0.005	**0.365 ± 0.0053**	0.3662 ± 0.0053	0.3694 ± 0.0065
GPI↓	0.0335 ± 0.0006	0.0273 ± 0.0018	0.0274 ± 0.0022	0.0268 ± 0.002	0.0261 ± 0.0018	**0.0257 ± 0.0019**	0.0262 ± 0.0019	0.0272 ± 0.0022
BHGI↑	0.7605 ± 0.0045	0.8085 ± 0.0133	0.8072 ± 0.0165	0.8114 ± 0.0151	0.817 ± 0.0138	**0.8195 ± 0.0147**	0.8158 ± 0.0148	0.8085 ± 0.0167
CI↓	0.0971 ± 0.0018	0.0800 ± 0.0040	0.0801 ± 0.0044	0.0788 ± 0.0037	0.0773 ± 0.0033	**0.0768 ± 0.0035**	0.0776 ± 0.0036	0.0796 ± 0.0043
TI↑	0.4025 ± 0.0026	0.4314 ± 0.008	0.4305 ± 0.0101	0.4327 ± 0.0093	0.4361 ± 0.0086	**0.4376 ± 0.0092**	0.4354 ± 0.0093	0.4312 ± 0.0104
DGI↑	0.0131 ± 0.0098	0.0113 ± 0.009	0.0214 ± 0.0044	0.0241 ± 0.0062	0.0241 ± 0.0071	0.0226 ± 0.0076	0.0245 ± 0.0076	**0.0268 ± 0.0089**
RLI↑	**0.3503 ± 0.0005**	0.3449 ± 0.0058	0.3458 ± 0.0079	0.3454 ± 0.0069	0.3433 ± 0.0068	0.3415 ± 0.0073	0.3438 ± 0.0072	0.3462 ± 0.0079
CHI↑	**580.2544 ± 6.2935**	514.6513 ± 64.2013	525.7927 ± 90.9895	524.7544 ± 71.9468	510.0968 ± 74.9872	489.8569 ± 80.0607	514.7695 ± 79.1941	532.1621 ± 91.2768
RTI↓	**0.5811 ± 0.0243**	0.6085 ± 0.0733	0.6181 ± 0.0941	0.6164 ± 0.0897	0.6284 ± 0.0854	0.659 ± 0.0904	0.6256 ± 0.0879	0.5999 ± 0.0959
WGI↑	0.482 ± 0.0036	0.4845 ± 0.013	0.4872 ± 0.0137	**0.4907 ± 0.0041**	0.4887 ± 0.0027	0.4895 ± 0.0022	0.4891 ± 0.0027	0.4882 ± 0.0099
DI↑	0.0015 ± 0.0011	0.0016 ± 0.0012	0.003 ± 0.0005	0.0033 ± 0.0006	0.0034 ± 0.0006	0.0032 ± 0.0007	0.0034 ± 0.0007	**0.0036 ± 0.0007**
TWBI↑	**5.9131 ± 0.0761**	5.6667 ± 0.6746	5.5636 ± 0.7249	5.5627 ± 0.6696	5.5409 ± 0.7499	5.2615 ± 0.77	5.5540 ± 0.7614	5.6614 ± 0.7409
BHI↑	0.0106 ± 0.0001	0.0152 ± 0.0027	0.0151 ± 0.0037	0.015 ± 0.0031	0.0158 ± 0.0032	**0.0167 ± 0.0034**	0.0156 ± 0.0034	0.015 ± 0.0038
PBMI↑	0.0218 ± 0.0007	0.0247 ± 0.0025	0.0238 ± 0.0014	0.0248 ± 0.0015	**0.0255 ± 0.0014**	0.025 ± 0.0011	0.0253 ± 0.0014	0.0241 ± 0.0015
XBI↓	21583.0483 ± 20106.9153	24404.8432 ± 36881.8857	1634.3854 ± 515.7399	1275.047 ± 432.8792	1149.9033 ± 297.29	1260.0305 ± 305.6108	1160.1052 ± 299.5600	**1116.3151 ± 306.0384**
DBI↓	**0.9033 ± 0.0083**	1.0726 ± 0.1325	1.0625 ± 0.1622	1.0501 ± 0.1291	1.1029 ± 0.1134	1.1265 ± 0.1212	1.0937 ± 0.1214	1.0563 ± 0.1485
DRI↑	**15.0016 ± 0.2518**	13.5595 ± 1.9698	13.7298 ± 2.8103	13.5771 ± 2.3868	13.1675 ± 2.5270	12.4065 ± 2.6831	13.2967 ± 2.6439	13.9822 ± 2.8417
KDWI↓	**977.5146 ± 16.282**	1104.4573 ± 170.5992	1113.2937 ± 235.4821	1114.6036 ± 210.4013	1153.9904 ± 224.468	1229.8482 ± 235.5594	1146.0096 ± 231.6428	1094.1964 ± 238.0447
LDRI↑	**2708.0265 ± 16.7196**	2596.6638 ± 149.8041	2598.9274 ± 210.2179	2592.8983 ± 183.0573	2559.9236 ± 194.6549	2497.4316 ± 205.5728	2568.2981 ± 202.2223	2616.9463 ± 212.5089
LSSRI↑	**1.0711 ± 0.0108**	0.9437 ± 0.1255	0.9582 ± 0.1748	0.9615 ± 0.1398	0.9321 ± 0.146	0.8899 ± 0.1554	0.9402 ± 0.1533	0.9704 ± 0.1743
TWI↓	**10.4405 ± 0.0842**	11.4881 ± 1.0373	11.3953 ± 1.4334	11.3491 ± 1.1519	11.5937 ± 1.2042	11.9514 ± 1.2803	11.5305 ± 1.2613	11.2956 ± 1.4253
U724_010	CNO-CC	TS [20]	GRASP-VND [20]	GRASP-VND-TS [20]	IVNS [21]	GVNS [25]	SGVNS [25]	NDVNS [26]
MRI↓	**0.2857 ± 0.0002**	0.2904 ± 0.0021	0.288 ± 0.002	0.2886 ± 0.0021	0.2869 ± 0.0015	0.2874 ± 0.0018	0.2872 ± 0.0016	0.2872 ± 0.0013
GPI↓	**0.0067 ± 0.0000**	0.0072 ± 0.0002	0.007 ± 0.0002	0.0071 ± 0.0003	0.0069 ± 0.0002	0.0069 ± 0.0002	0.0069 ± 0.0002	0.0069 ± 0.0002
BHGI↑	**0.9256 ± 0.0003**	0.9189 ± 0.0025	0.9218 ± 0.0026	0.9208 ± 0.0028	0.923 ± 0.002	0.9225 ± 0.0024	0.9228 ± 0.0021	0.9228 ± 0.0018
CI↓	**0.0431 ± 0.0002**	0.0468 ± 0.0013	0.0452 ± 0.0013	0.0457 ± 0.0013	0.0446 ± 0.0010	0.0449 ± 0.0011	0.0447 ± 0.0010	0.0447 ± 0.0009
TI↑	**0.3928 ± 0.0007**	0.3883 ± 0.001	0.3893 ± 0.0011	0.389 ± 0.0011	0.3899 ± 0.0009	0.3897 ± 0.0009	0.3898 ± 0.0009	0.3898 ± 0.0008
DGI↑	**0.1575 ± 0.0397**	0.104 ± 0.0372	0.1413 ± 0.0408	0.1362 ± 0.0437	0.1287 ± 0.0273	0.1526 ± 0.0306	0.1429 ± 0.0299	0.1255 ± 0.0277
RLI↑	**0.3009 ± 0.0000**	0.3002 ± 0.0003	0.3005 ± 0.0003	0.3004 ± 0.0003	0.3006 ± 0.0002	0.3005 ± 0.0003	0.3006 ± 0.0002	0.3006 ± 0.0002
CHI↑	**758.4242 ± 0.5499**	722.6965 ± 12.9587	737.7558 ± 14.1883	732.61 ± 15.2223	743.5445 ± 9.5301	741.2396 ± 11.9724	743.0472 ± 10.0315	742.7902 ± 9.0852
RTI↓	**0.242 ± 0.0039**	0.2926 ± 0.0237	0.2762 ± 0.0315	0.2857 ± 0.0356	0.2629 ± 0.0202	0.2685 ± 0.0225	0.2734 ± 0.0241	0.264 ± 0.0196
WGI↑	**0.4957 ± 0.0045**	0.4804 ± 0.0084	0.4851 ± 0.0047	0.4848 ± 0.0047	0.4878 ± 0.0052	0.4875 ± 0.0059	0.4867 ± 0.0073	0.4855 ± 0.0056
DI↑	**0.0311 ± 0.0083**	0.0195 ± 0.0064	0.0276 ± 0.0084	0.0257 ± 0.0088	0.0247 ± 0.0057	0.0289 ± 0.0053	0.0271 ± 0.0049	0.0238 ± 0.0051
TWBI↑	**19.1542 ± 0.0452**	18.5859 ± 0.3061	18.9210 ± 0.2727	18.7672 ± 0.3073	18.9646 ± 0.2026	18.9333 ± 0.2865	19.0032 ± 0.2457	19.0085 ± 0.1951
BHI↑	0.0164 ± 0.0001	**0.0173 ± 0.0003**	0.0169 ± 0.0003	0.017 ± 0.0003	0.0168 ± 0.0002	0.0169 ± 0.0002	0.0168 ± 0.0002	0.0168 ± 0.0002
PBMI↑	**0.1095 ± 0.0007**	0.1043 ± 0.003	0.1068 ± 0.0041	0.1063 ± 0.0034	0.1061 ± 0.0036	0.1068 ± 0.0033	0.1071 ± 0.0031	0.1067 ± 0.0028
XBI↓	102.8882 ± 68.3431	257.3993 ± 204.5996	129.7696 ± 75.3984	145.1907 ± 80.1117	137.9924 ± 62.9588	**98.2766 ± 46.4781**	113.7271 ± 56.2674	145.1594 ± 63.6229
DBI↓	**0.8216 ± 0.0124**	0.8708 ± 0.0233	0.861 ± 0.0158	0.8677 ± 0.0201	0.8562 ± 0.0204	0.855 ± 0.0187	0.8545 ± 0.0200	0.8587 ± 0.0160
DRI↑	**111.3492 ± 0.4285**	103.4431 ± 2.7301	107.0575 ± 3.0396	105.6642 ± 3.6039	108.0547 ± 1.9102	107.536 ± 2.7541	108.1181 ± 2.0899	108.2695 ± 1.6049
KDWI↓	**3524.9234 ± 13.5417**	3796.7629 ± 99.1802	3669.0281 ± 105.7227	3718.6638 ± 127.8848	3633.4428 ± 65.3669	3652.2013 ± 95.9685	3631.5602 ± 73.2004	3625.9105 ± 54.7421
LDRI↑	**3411.969 ± 2.7839**	3358.4141 ± 19.0098	3383.2373 ± 20.7102	3373.6303 ± 24.7978	3390.1208 ± 12.9117	3386.516 ± 18.7835	3390.5219 ± 14.2905	3391.5907 ± 10.8306
LSSRI↑	**2.2576 ± 0.0007**	2.2092 ± 0.0179	2.2298 ± 0.0193	2.2227 ± 0.0208	2.2377 ± 0.0128	2.2345 ± 0.0162	2.2370 ± 0.0136	2.2367 ± 0.0122
TWI↓	**11.9031 ± 0.0078**	12.4365 ± 0.2002	12.2078 ± 0.2138	12.2857 ± 0.2314	12.1199 ± 0.1409	12.1549 ± 0.1792	12.1274 ± 0.1497	12.1309 ± 0.1343

nections in a fully connected data graph) are quadratically proportional to the cluster sizes. To remedy the shortcoming, the quadratic assignment problem formulation can be normalized by using cluster cardinality. The normalized total within-cluster distance for the circled clusters in Fig. 1 is $1 + \sqrt{2}/2$ ($\approx$1.707), much smaller than that of the unnatural clusters surrounded by the ellipses ($(2 + \sqrt{2})/3 + 3/2 \approx 2.638$).

Let n_l be the number of data points in cluster l for $l = 1, 2, \ldots, p$. As such, the quadratic assignment problem for capacitated clustering is reformulated as follows:

$$\begin{aligned} \min_{x} \quad & \sum_{l=1}^{p} \frac{\sum_{i=1}^{n} \sum_{j<i} d_{ij} x_{il} x_{jl}}{\sum_{i=1}^{n} x_{il}} \\ \text{s.t.} \quad & \sum_{l=1}^{p} x_{il} = 1, \quad i = 1, 2, \ldots, n \\ & \sum_{i=1}^{n} a_{il} x_{il} \leq b_l, \quad l = 1, 2, \ldots, p \\ & x_{il} \in \{0, 1\}, \quad i = 1, 2, \ldots, n;\ l = 1, 2, \ldots, p. \end{aligned} \tag{8}$$

TABLE V

MEAN VALUES AND STANDARD DEVIATIONS OF 21 INTERNAL CLUSTER VALIDITY CRITERIA RESULTING FROM CNO-CC AND SEVEN BASELINES ON RL1304_010 ($n = 1304,\ p = 10$, AND $b = 7237$) AND DONI2 ($n = 2000,\ p = 6$, AND $b = 400$)

Rl1304_010	CNO-CC	TS [20]	GRASP-VND [20]	GRASP-VND-TS [20]	IVNS [21]	GVNS [25]	SGVNS [25]	NDVNS [26]
MRI↓	**0.2961 ± 0.0017**	0.2994 ± 0.0028	0.3005 ± 0.0037	0.2987 ± 0.0030	0.2982 ± 0.0031	0.2975 ± 0.0031	0.2967 ± 0.0022	0.2977 ± 0.0038
GPI↓	**0.0082 ± 0.0002**	0.0086 ± 0.0003	0.0087 ± 0.0005	0.0085 ± 0.0004	0.0084 ± 0.0004	0.0083 ± 0.0003	0.0082 ± 0.0003	0.0084 ± 0.0005
BHGI↑	**0.9091 ± 0.0024**	0.9048 ± 0.0038	0.9031 ± 0.0052	0.9055 ± 0.0043	0.9064 ± 0.0044	0.9076 ± 0.0038	0.9084 ± 0.0029	0.907 ± 0.0053
CI↓	**0.0494 ± 0.0011**	0.0515 ± 0.0018	0.0522 ± 0.0023	0.0511 ± 0.0019	0.0507 ± 0.0019	0.0503 ± 0.002	0.0498 ± 0.0014	0.0504 ± 0.0024
TI↑	**0.3855 ± 0.0011**	0.3836 ± 0.0016	0.3829 ± 0.0022	0.384 ± 0.0018	0.3843 ± 0.0018	0.3847 ± 0.0016	0.3852 ± 0.0012	0.3846 ± 0.0022
DGI↑	0.0362 ± 0.0033	0.0345 ± 0.0007	0.0364 ± 0.0044	0.0366 ± 0.0065	0.0368 ± 0.0053	0.0364 ± 0.0043	0.0371 ± 0.004	**0.0386 ± 0.0064**
RLI↑	0.2984 ± 0.0004	0.2979 ± 0.0008	0.2977 ± 0.0008	0.298 ± 0.0007	0.2981 ± 0.0009	0.2983 ± 0.0005	**0.2984 ± 0.0005**	0.2983 ± 0.0006
CHI↑	**1224.9948 ± 19.9023**	1189.2908 ± 33.4863	1175.1102 ± 43.5504	1193.706 ± 37.9317	1202.5524 ± 38.5758	1212.973 ± 30.3621	1219.3629 ± 24.8813	1207.0595 ± 45.7817
RTI↓	0.3207 ± 0.0360	0.3272 ± 0.0424	0.3312 ± 0.0312	0.3258 ± 0.0562	**0.3111 ± 0.0462**	0.3143 ± 0.0338	0.3265 ± 0.0348	0.3215 ± 0.0400
WGI↑	**0.4786 ± 0.0085**	0.4687 ± 0.0089	0.4672 ± 0.0084	0.4701 ± 0.0100	0.4737 ± 0.0078	0.4741 ± 0.0113	0.4773 ± 0.0074	0.4734 ± 0.0104
DI↑	0.0075 ± 0.0009	0.007 ± 0.0005	0.0074 ± 0.0011	0.0075 ± 0.0013	0.0076 ± 0.0011	0.0076 ± 0.0008	0.0078 ± 0.0011	**0.008 ± 0.0012**
TWBI↑	**17.6579 ± 0.6551**	17.4335 ± 0.9964	17.2613 ± 1.0668	17.2654 ± 1.0337	17.423 ± 0.9258	17.3198 ± 0.6741	17.4921 ± 0.695	17.2546 ± 0.795
BHI↑	0.0164 ± 0.0002	0.0168 ± 0.0005	**0.017 ± 0.0006**	0.0168 ± 0.0005	0.0167 ± 0.0005	0.0165 ± 0.0004	0.0164 ± 0.0003	0.0166 ± 0.0006
PBMI↑	**0.0899 ± 0.0025**	0.0870 ± 0.0027	0.0854 ± 0.0038	0.0868 ± 0.0033	0.0868 ± 0.0027	0.0882 ± 0.0033	0.0891 ± 0.0025	0.0872 ± 0.0052
XBI↓	1325.6158 ± 184.5429	1439.6387 ± 36.9777	1333.4594 ± 242.7993	1344.7535 ± 271.0738	1318.0585 ± 255.9138	1322.0595 ± 220.8799	1282.3469 ± 233.7296	**1215.1771 ± 306.4138**
DBI↓	**0.865 ± 0.0187**	0.8909 ± 0.0307	0.9032 ± 0.0296	0.8874 ± 0.0356	0.8782 ± 0.0339	0.8705 ± 0.0261	0.8733 ± 0.0223	0.8759 ± 0.0356
DRI↑	**89.5993 ± 2.6320**	86.0073 ± 3.7855	84.4232 ± 4.9749	85.9522 ± 4.9607	87.2963 ± 3.9081	87.7477 ± 3.9199	88.7901 ± 3.1618	87.3838 ± 4.7844
KDWI↓	**10930.8651 ± 327.2579**	11399.1741 ± 509.9653	11630.5094 ± 700.8956	11422.4393 ± 680.9838	11232.5939 ± 535.0002	11173.8979 ± 511.5732	11034.902 ± 402.9685	11233.5019 ± 671.6296
LDRI↑	**5861.3923 ± 38.6691**	5807.3669 ± 57.8638	5782.1566 ± 77.6824	5805.6358 ± 76.4867	5826.6846 ± 60.1948	5833.4468 ± 58.9885	5849.3037 ± 47.0158	5827.3172 ± 74.5579
LSSRI↑	**2.1423 ± 0.0163**	2.1125 ± 0.0285	2.1002 ± 0.0372	2.1161 ± 0.0322	2.1234 ± 0.0327	2.1323 ± 0.0253	2.1376 ± 0.0206	2.127 ± 0.0393
TWI↓	**21.101 ± 0.3098**	21.6750 ± 0.5567	21.9175 ± 0.729	21.6072 ± 0.6282	21.4657 ± 0.6350	21.2939 ± 0.4844	21.1907 ± 0.3944	21.4017 ± 0.7728
Doni2	CNO-CC	TS [20]	GRASP-VND [20]	GRASP-VND-TS [20]	IVNS [21]	GVNS [25]	SGVNS [25]	NDVNS [26]
MRI↓	0.3753 ± 0.0118	0.6574 ± 0.1238	0.3704 ± 0.0046	0.3738 ± 0.0086	0.3703 ± 0.0047	0.3674 ± 0.0058	**0.3672 ± 0.0064**	0.3705 ± 0.0050
GPI↓	0.0305 ± 0.0025	0.0859 ± 0.0242	0.0289 ± 0.0014	0.0300 ± 0.0020	0.0288 ± 0.0014	**0.0279 ± 0.0014**	0.0281 ± 0.0019	0.0290 ± 0.0015
BHGI↑	0.786 ± 0.0188	0.3854 ± 0.1755	0.7978 ± 0.0112	0.7902 ± 0.0147	0.7988 ± 0.0115	**0.8062 ± 0.0112**	0.8050 ± 0.0142	0.7972 ± 0.0122
CI↓	0.0842 ± 0.0069	0.237 ± 0.0673	0.0812 ± 0.0031	0.0832 ± 0.0051	0.0811 ± 0.0031	0.0791 ± 0.0038	**0.0790 ± 0.0041**	0.0812 ± 0.0032
TI↑	0.4194 ± 0.0111	0.2045 ± 0.0951	0.4269 ± 0.0076	0.4224 ± 0.0085	0.4275 ± 0.0079	**0.4329 ± 0.0082**	0.4322 ± 0.0096	0.4268 ± 0.0082
DGI↑	**0.0118 ± 0.0054**	0.0014 ± 0.0037	0.0104 ± 0.0049	0.0106 ± 0.0041	0.0109 ± 0.0049	0.0064 ± 0.0041	0.0094 ± 0.0044	0.0099 ± 0.0049
RLI↑	**0.3452 ± 0.0034**	0.2698 ± 0.033	0.3401 ± 0.0069	0.3423 ± 0.0052	0.3393 ± 0.0058	0.3350 ± 0.0064	0.3368 ± 0.0062	0.3408 ± 0.0074
CHI↑	**1004.364 ± 70.9312**	372.9465 ± 278.6564	918.6524 ± 121.5407	950.9359 ± 85.3799	900.8383 ± 93.1683	832.1556 ± 102.1302	858.3044 ± 99.4213	926.1673 ± 123.5089
RTI↓	**0.7277 ± 0.0926**	286.6146 ± 756.3596	0.783 ± 0.1265	0.7693 ± 0.0926	0.8277 ± 0.0832	0.8369 ± 0.1179	0.8218 ± 0.1247	0.7578 ± 0.1395
WGI↑	0.4805 ± 0.0075	0.1282 ± 0.1523	**0.4819 ± 0.0062**	0.4791 ± 0.0161	0.4809 ± 0.0058	0.4760 ± 0.0126	0.4771 ± 0.0105	0.4770 ± 0.0068
DI↑	**0.0018 ± 0.0009**	0.0002 ± 0.0006	0.0017 ± 0.0007	0.0016 ± 0.0006	0.0017 ± 0.0008	0.0011 ± 0.0007	0.0016 ± 0.0007	0.0016 ± 0.0008
TWBI↑	**5.2979 ± 0.2341**	2.0885 ± 1.5435	4.9342 ± 0.6000	5.1243 ± 0.3938	4.9149 ± 0.5014	4.5688 ± 0.5101	4.7114 ± 0.4733	4.9561 ± 0.5253
BHI↑	0.0159 ± 0.0016	**0.0259 ± 0.0041**	0.0185 ± 0.0039	0.0173 ± 0.0023	0.0189 ± 0.0035	0.0219 ± 0.0043	0.0212 ± 0.0044	0.0186 ± 0.0039
PBMI↑	0.0247 ± 0.0015	0.0088 ± 0.0078	0.0254 ± 0.0012	0.0249 ± 0.0018	0.0255 ± 0.0013	0.0261 ± 0.0018	**0.0263 ± 0.0017**	0.0254 ± 0.0015
XBI↓	9298.6004 ± 9836.2687	Inf ± NaN	**9057.8492 ± 9446.3455**	9366.0864 ± 12640.6587	9209.1882 ± 11820.1008	39882.496 ± 59725.9982	11666.0136 ± 13644.2871	17788.0306 ± 36524.9046
DBI↓	**0.9712 ± 0.0435**	8.3499 ± 7.7375	1.0194 ± 0.0646	1.0004 ± 0.0511	1.0368 ± 0.0623	1.0795 ± 0.0693	1.0692 ± 0.0694	1.0165 ± 0.0801
DRI↑	**12.8707 ± 1.0474**	4.4601 ± 3.9682	11.5566 ± 1.9801	12.1153 ± 1.2931	11.3390 ± 1.4734	10.2417 ± 1.6327	10.6743 ± 1.5576	11.6810 ± 1.8693
KDWI↓	**5192.2762 ± 388.936**	20608.2012 ± 6833.5656	5918.6766 ± 1050.4762	5553.2343 ± 690.4166	5966.2302 ± 865.0884	6635.1068 ± 979.3266	6350.7247 ± 912.4482	5840.9621 ± 1016.3060
LDRI↑	**5103.9472 ± 156.2614**	2559.2783 ± 1148.3803	4865.817 ± 350.2494	4977.0541 ± 229.3805	4839.1239 ± 275.8605	4629.8417 ± 308.3246	4715.4927 ± 291.2164	4890.1369 ± 334.5635
LSSRI↑	**0.9214 ± 0.0683**	-0.2262 ± 0.5026	0.8261 ± 0.1333	0.865 ± 0.0935	0.8095 ± 0.1072	0.7287 ± 0.1195	0.7602 ± 0.1153	0.8338 ± 0.1372
TWI↓	**24.5314 ± 1.1795**	48.002 ± 10.2581	26.2797 ± 2.4352	25.5457 ± 1.7064	26.5598 ± 1.9908	28.0746 ± 2.2330	27.4797 ± 2.1479	26.1414 ± 2.5181

Note that $\sum_{i=1}^{n} x_{il} = n_l$. In addition, if $d_{ij} = d_{ji}$ and $d_{ii} = 0$, then $\sum_{i=1}^{n} \sum_{j<i} d_{ij} x_{il} x_{jl} = \sum_{i=1}^{n} \sum_{j=1}^{n} d_{ij} x_{il} x_{jl}/2$.

If $d_{ij} = \|\underline{v}_i - \underline{v}_j\|$, then we have

$$\begin{aligned}
&\sum_{l=1}^{p} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} d_{ij} x_{il} x_{jl}}{\sum_{j=1}^{n} x_{jl}} \\
&= \sum_{l=1}^{p} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} \left\|\underline{v}_i - \underline{v}_j\right\| x_{il} x_{jl}}{\sum_{j=1}^{n} x_{jl}} \\
&\geq \sum_{l=1}^{p} \frac{\sum_{i=1}^{n} \left\|\sum_{j=1}^{n} \left(\underline{v}_i - \underline{v}_j\right) x_{jl}\right\| x_{il}}{\sum_{j=1}^{n} x_{jl}} \\
&= \sum_{l=1}^{p} \sum_{i=1}^{n} \left\| \frac{\sum_{j=1}^{n} \left(\underline{v}_i - \underline{v}_j\right) x_{jl}}{\sum_{j=1}^{n} x_{jl}} \right\| x_{il} \\
&= \sum_{l=1}^{p} \sum_{i=1}^{n} \left\| \underline{v}_i - \frac{\sum_{j=1}^{n} x_{jl} \underline{v}_j}{\sum_{j=1}^{n} x_{jl}} \right\| x_{il}
\end{aligned}$$

where $\sum_{j=1}^{n} x_{jl}\underline{v}_j / \sum_{j=1}^{n} x_{jl}$ is the centroid of cluster l. It implies that the objective function (8) is an upper bounding function of the objective function in k-means clustering. Nevertheless, the dissimilarity coefficient d_{ij} in (8) is more general beyond the definition of any norm.

In view that the minimization of a fractional objective function in (8) can be carried out via the minimization of its numerator and the maximization of its denomination simultaneously, problem (8) is reformulated as the following mixed-integer quadratic programming problem:

$$\begin{aligned}
\min_{x,\lambda}\ & \sum_{l=1}^{p} \sum_{i=1}^{n} \left(\sum_{j<i} d_{ij} x_{il} x_{jl} - \lambda_l x_{il} \right) \\
\text{s.t.}\ & \sum_{l=1}^{p} x_{il} = 1, \quad i = 1, 2, \ldots, n \\
& \sum_{i=1}^{n} a_{il} x_{il} \leq b_l, \quad l = 1, 2, \ldots, p \\
& x_{il} \in \{0, 1\}, \quad i = 1, 2, \ldots, n;\ l = 1, 2, \ldots, p
\end{aligned} \tag{9}$$

where λ_l is a positive of weight for $l = 1, 2, \ldots, p$. Let $\phi_\lambda(X)$ denote the objective function to be minimized in problem (9) hereafter.

For given nonzero initial solution $x(0)$, problem (9) can be solved with iterative weights as follows: For $t = 0, 1, \ldots$;

$$\lambda_l(t) = \frac{\sum_{i=1}^{n} \sum_{j<i} d_{ij} x_{il}(t) x_{jl}(t)}{\sum_{i=1}^{n} x_{il}(t)}, \quad l = 1, 2, \ldots, p \tag{10a}$$

$$x(t+1) = \arg\min_{x \in \mathcal{X}} \sum_{l=1}^{p} \sum_{i=1}^{n} \left(\sum_{j<i} d_{ij} x_{il} x_{jl} - \lambda_l(t) x_{il} \right) \tag{10b}$$

where $\mathcal{X} = \{x \in \{0,1\}^{n \times p} | \sum_{l=1}^{p} x_{il} = 1, \sum_{i=1}^{n} a_{il} x_{il} \leq b_l, i = 1, 2, \ldots, n; l = 1, 2, \ldots, p\}$. Problem (9) can then be solved as an iteratively reweighted binary quadratic programming problem for given $\lambda_l(t)$ $(l = 1, 2, \ldots, p)$ updated iteratively according to (10a). Because of the effect of minimization in (9)

$$\lambda_l(t+1) \geq \lambda_l(t) = \frac{\sum_{i=1}^{n} \sum_{j<i} d_{ij} x_{il}(t) x_{jl}(t)}{\sum_{i=1}^{n} x_{il}(t)}.$$

The above-mentioned inequality implies that the objective function in (9) is a surrogate function in a majorization-minimization framework [81], [82], [83], [84]. It is proven that the convergence rate of the weight λ in the iterative reweighting procedure (10) is quadratic [85].

The capacitated clustering can be further reformulated in an iteratively reweighted quadratic unconstrained binary optimization problem by introducing a quadratic penalty function into the objective function as an alternative to imposing constraints. The terms in the penalty function are chosen so that the influence of the original constraints on the solution process can alternatively be enforced by the natural functioning of the optimizer as it looks for solutions that avoid incurring the penalties. That is, the penalty terms are formulated so that they equal zero for feasible solutions and are positive for infeasible solutions. For a minimization problem, these penalty functions are added to create an augmented objective function to be minimized. If all the penalty terms can be driven to zero, the augmented objective function becomes the original objective function to be minimized.

To handle the equality constraints in (9), a quadratic penalty term is defined as follows:

$$\phi_1(X) = \frac{1}{2}\sum_{i=1}^{n}\left(\sum_{l=1}^{p} x_{il} - 1\right)^2. \tag{11}$$

In view that $x_{il} \in \{0,1\}$, and hence, $x_{il}^2 = x_{il}$, the penalty term can be equivalently expressed as follows:

$$\phi_1(X) = \frac{1}{2}\sum_{i=1}^{n}\left(\sum_{l=1}^{p} x_{il} - 2\sum_{l=1}^{p}\sum_{j<l} x_{ij}x_{il} - 1\right). \tag{12}$$

A penalty term for the inequality constraint in (13) by using rectified activation function is defined as

$$\phi_2(X) = \frac{1}{2}\sum_{l=1}^{p}\left(\max\left\{0, \sum_{i=1}^{n} a_{il}x_{il} - b_l\right\}\right)^2. \tag{13}$$

With the two penalty functional terms and the quadratic objective function to be minimized, a penalty function and a penalized or augmented objective function are defined as follows:

$$\phi(X) = \phi_1(X) + \phi_2(X) \tag{14}$$

$$f_\lambda(X) = \phi_\lambda(X) + \rho\phi(X) \tag{15}$$

where ρ is a positive penalty parameter, and $X \in \{0,1\}^{n\times p}$.

Based on the penalty function (14) and penalized objective function in (15), problem (9) is reformulated as the following quadratic unconstrained binary optimization problem:

$$\min_X \; f_\lambda(X), \quad \text{s.t. } X \in \{0,1\}^{n\times p}. \tag{16}$$

It is known that problems (16) and (9) are equivalent in terms of their optimal solutions if the penalty parameter is sufficiently large [86].

IV. Algorithm Description

Based on the formulated problem in (16), a CNO-driven capacitated clustering algorithm termed CNO-CC is developed.

In analogy to DHNm, a BM with a momentum term (BMm) is defined as follows:

$$\begin{cases} u(t+1) = u(t) + Wx(t) - \theta \\ p(x_i(t)=1) = \dfrac{1}{1+\exp\left(-\frac{u_i(t)}{T}\right)} \\ p(x_i(t)=0) = 1 - p(x_i(t)=1). \end{cases} \tag{17}$$

In the CNO-CC algorithm, a population of BMms in (17) are employed for distributed local search, and the PSO rule in (6) is used to repetitively update DHN or BM neuronal states upon their local convergence to escape from local optima and move toward global optimal solutions.

Algorithm 1 CNO-CC

Input: Dissimilarity coefficient matrix D, DHN or BM population size N, termination criterion M, PSO parameters c_0, c_1, and c_2, initial temperature parameter T_0.

```
1  For i = 1, 2, ..., N, generate random initial neuronal
   state matrices X_i(0) ∈ {0,1}^{n×p}, PSO velocity matrices
   V_i ∈ [−1,1]^{n×p},; set initial group-best matrix and initial
   individual-best matrices for the ith BM X* = X_i* = 0.
   Set iteration counter t = 0 and termination counter ℓ = 0;
2  while ℓ < M do
3      for i = 1 to N do
4          flg_i = 0;
5          while flg_i = 0 do
6              Update X(t)_i according to (1) for DHN or (4)
               for BM;
7              if X_i(t) = X_i(t+1) then
8                  Compute λ_i^l (l = 1, ..., p) according
                   to (10a);
9                  flg_i = 1;
10             end
11             t ← t + 1;
12         end
13         if f_ρ(X_i(t)) < f_ρ(X_i*) then
14             X_i* ← X_i(t);
15         end
16     end
17     if min{f_λ(X_1*), ..., f_λ(X_N*)} < f_λ(X*) then
18         X* ← arg min{f_λ(X_1*), ..., f_λ(X_N*)};
19         ℓ ← 0;
20     else
21         ℓ ← ℓ + 1;
22     end
23     for i = 1 to N do
24         Update X_i(t+1) according to (6);
25     end
26 end
```

Output: X^*.

Algorithm 1 describes the CNO-CC algorithm. Steps 3–12 are to obtain the equilibria of BMms for scatter local searches.

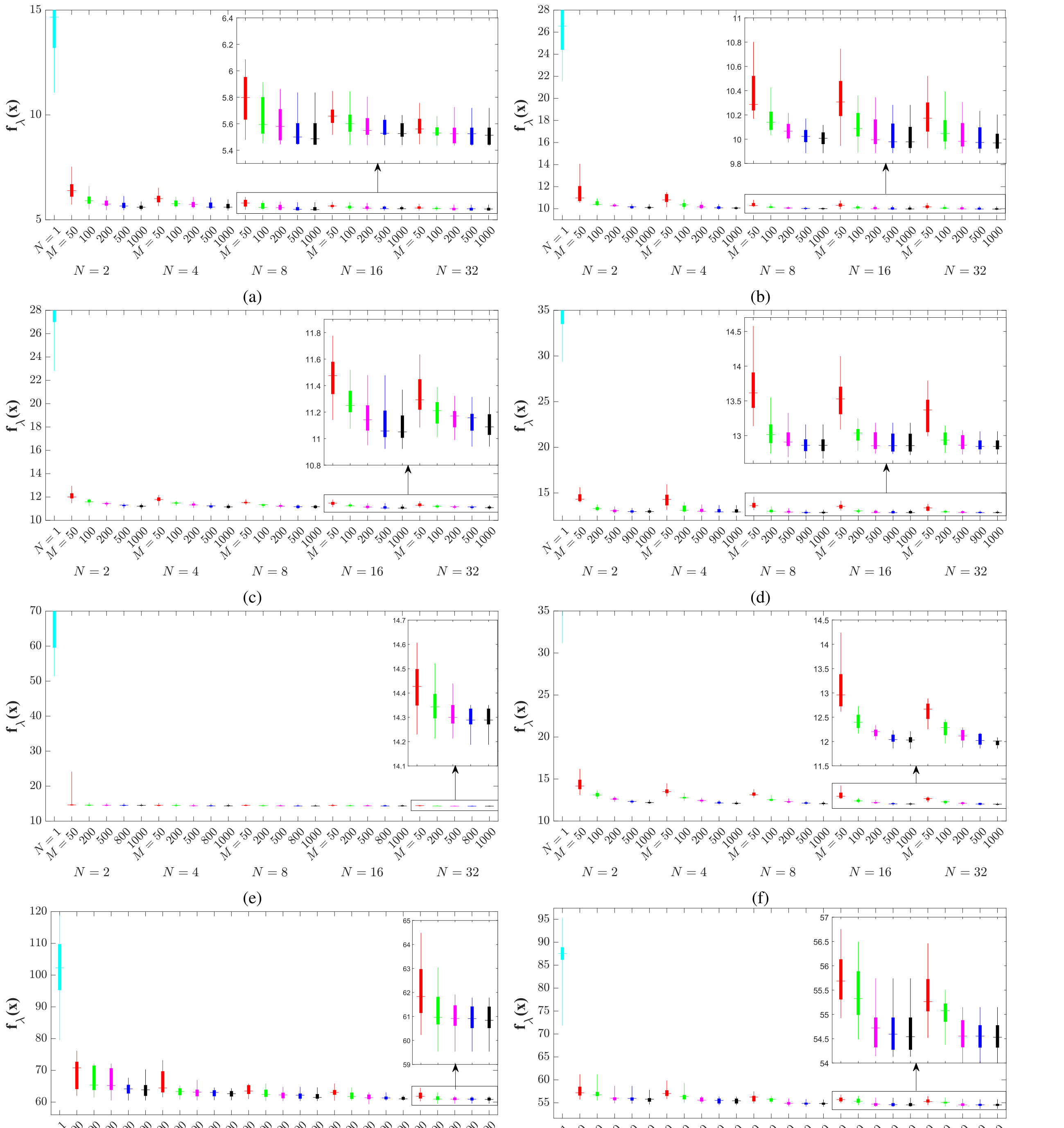

Fig. 2. Monte Carlo test results of the CNO-CC algorithm with several values of M and N on the ten datasets. (a) SJC1. (b) SJC2. (c) SJC3a. (d) SJC3b. (e) SJC4a. (f) SJC4b. (g) U724_010. (h) Doni1.

Steps 13–14 and 17–18 are to update individual and population best solutions, respectively. Steps 23–25 are to reposition the neuronal states using the PSO rule.

In the CNO-CC algorithm, there are two important hyperparameters: the BMm population size denoted by N and the minimum number of consecutive iterations without further improvement denoted by M as the termination criterion. These two hyperparameters usually need to be determined in an ad hoc manner, as they depend on the inherent complexity of the problem under study and the desired spatial and temporal complexities of the solution method. In general, the larger the population size N is, the faster CNO-CC converges to optimal or acceptable high-quality solutions. Therefore, it is a trade-off between the spatial and temporal complexities. Due to the

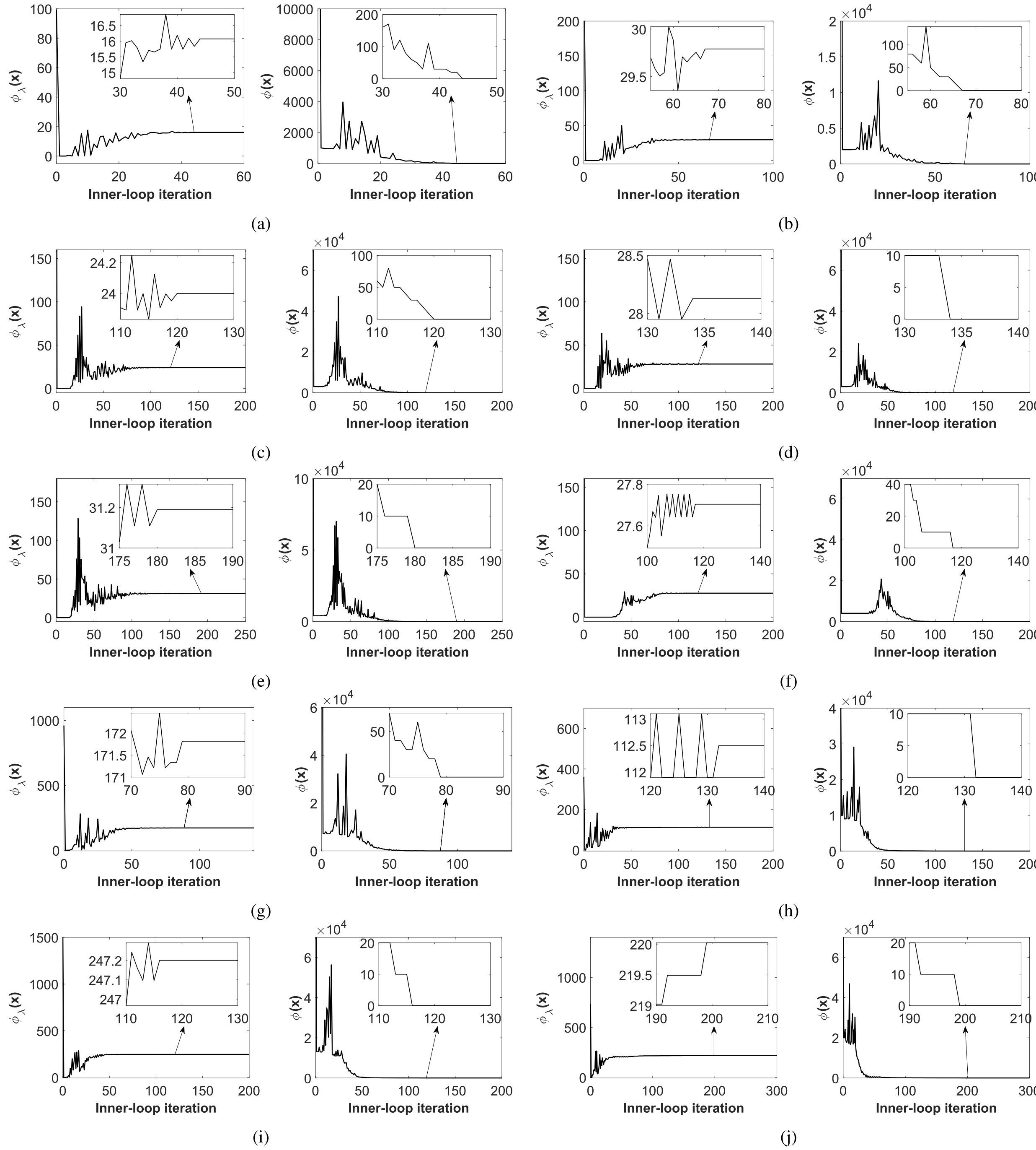

Fig. 3. Snapshots of objective function value and penalty function value of the CNO-CC algorithm. (a) SJC1. (b) SJC2. (c) SJC3a. (d) SJC3b. (e) SJC4a. (f) SJC4b. (g) U724_010. (h) Doni1. (i) Rl1304_010. (j) Doni2.

stochastic nature of PSO-based reinitialization for multistart scatter search, M needs to be set with a fair value to reach the theoretically proven almost-sure convergence.

V. Experimental Results

A. Experiment Setups

As the evaluation of clustering results is subjective, many cluster validity criteria are used to evaluate the multifaced goodness of clustering results. Internal cluster validity criteria are independent of the specific use of similarity or dissimilarity coefficients, whereas external cluster validity criteria depend on the similarity or dissimilarity coefficients used as well as labels. In this study, 21 label-free internal cluster validity criteria in Table I are used for evaluating the clustering performance.

The experiments are based on ten benchmark datasets with given data weights and cluster capacities (i.e., a_{il} and b_l in (7)), exclusively used for capacitated clustering: SJC1, SJC2, SJC3a, SJC3b, SJC4a, SJC4b[1] (used in [12], [13], [14], [15], [22], [87], [88], [89], [90], [91], and [92]), U724_010[2]

[1] http://www.lac.inpe.br/~lorena/instancias.html
[2] https://github.com/emuritiba/cccp

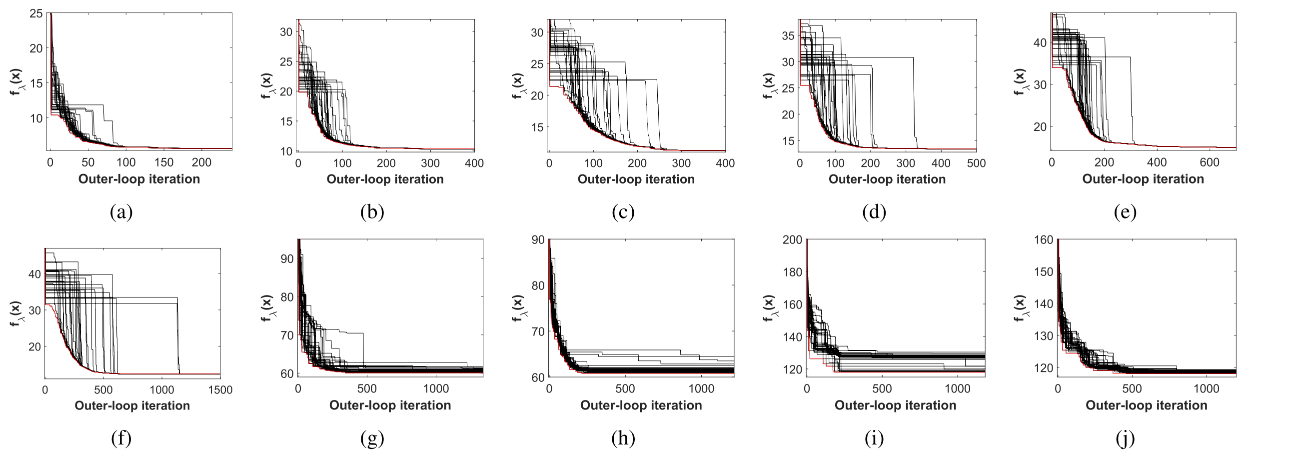

Fig. 4. Convergent behavior of the CNO-CC algorithm on the ten datasets with $M = 1000$ and $N = 32$. (a) SJC1. (b) SJC2. (c) SJC3a. (d) SJC3b. (e) SJC4a. (f) SJC4b. (g) U724_010. (h) Doni1. (i) Rl1304_010. (j) Doni2.

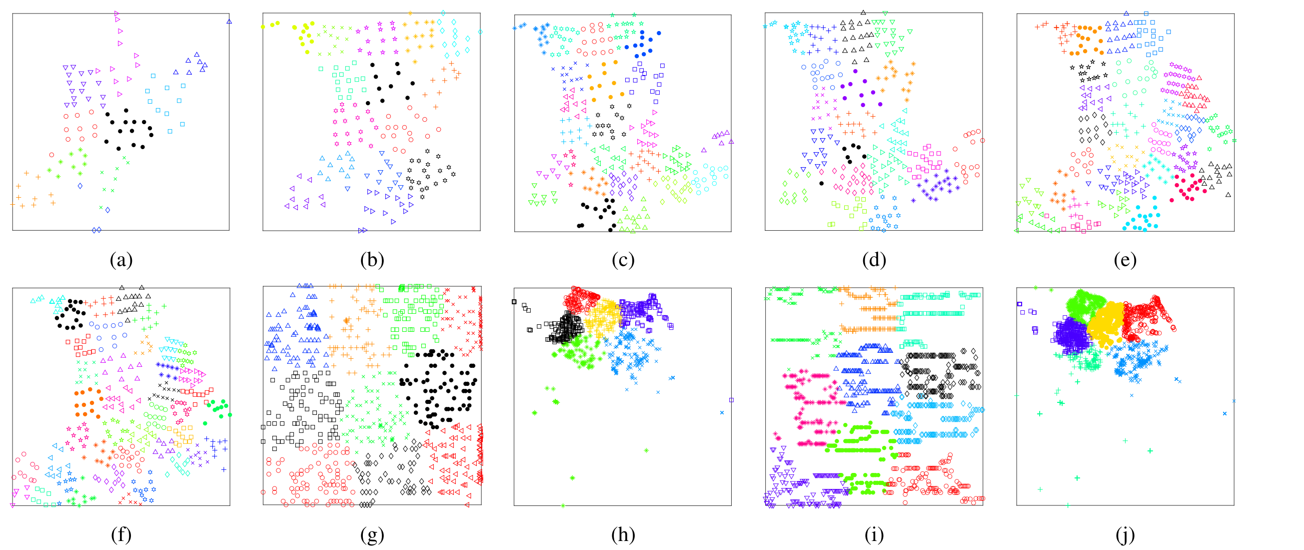

Fig. 5. Resulting clusters using the CNO-CC algorithm. (a) SJC1. (b) SJC2. (c) SJC3a. (d) SJC3b. (e) SJC4a. (f) SJC4b. (g) U724_010. (h) Doni1. (i) Rl1304_010. (j) Doni2.

(used in [22], [89], [91], and [92]), Doni1, Doni2[1] (used in [22], [87], [88], [90], [91], and [92]), and Rl1304_010[2] (used in [22], [89], and [91]), with their major parameters listed in Table II and the coordinates of the data points available for evaluating the internal validity criteria of clustering results. In this study, the Euclidean distance is used to represent the dissimilarity between data.

The proposed CNO-CC algorithm is compared with seven prevailing capacitated clustering algorithms: TS [20], greedy randomized adaptive search procedure with variable neighborhood descent (GRASP-VND) [20], GRASP-VND and TS (GRASP-VND-TS) [20], IVNS [21], GVNS [25], SGVNS [25], and NDVNS [26] algorithms. The code of CNO-CC is publicly accessible at Github.[3] The codes of the GRASP-VND, TS, GRASP-VND-TS, and IVNS algorithms are obtained following a link in [21]. The codes of GVNS and SGVNS are from http://www.mi.sanu.ac.rs/~nenad/ccp/ maintained by Brimberg et al. [25]. The codes of NDVNS are obtained following a link in [26].

[3]https://github.com/HongzongLI-CS/CNO-CC-Github

B. Parameters Selection

The values of two hyperparameters N (population size) and M (termination criteria) in Algorithm 1 are set based on 20-run Monte Carlo tests with random initial states on the ten datasets. Fig. 2 depicts the box-plots of the Monte Carlo test results obtained using the CNO-CC algorithm over 20 runs with different initial states on the ten datasets, where a center bar in a box marks the median, the top and bottom of the box denote the upper quartile $q_n(0.75)$ and the lower quartile $q_n(0.25)$, and the whiskers denote the highest or lowest values. As shown in Fig. 2, the objective function values reach their minimum in most runs with $M = 500$ and $N = 8$ on SJC1, $M = 500$ and $N = 8$ on SJC2, $M = 500$ and $N = 16$ on SJC3a, $M = 900$ and $N = 8$ on SJC3b, $M = 800$ and $N = 32$ on SJC4a, $M = 500$ and $N = 16$ on SJC4b, $M = 600$ and $N = 32$ on U724_010, and $M = 800$ and $N = 16$ on Doni1. Table II lists the values of the hyperparameters as well as other parameters used in the experiments. In the BMm, the cooling rate $\alpha = 0.2$. In the PSO update rule (6), $c_0 = 1$ and $c_1 = c_2 = 2$.

C. Neurodynamic Behaviors

Fig. 3 depicts ten snapshots of the convergent behaviors of the objective function $f(x)$ in (9) and the penalty function $p(x)$ in (14), resulting from an individual BMm in the inner-loop of CNO-CC (Step 6) on the ten datasets. The snapshots in Fig. 3 show that the objective function values increase because the solution does not satisfy the constraints, as shown in the penalty function values on the right-hand side. The objective functions reach stationary points, and the penalty function decreases to zero (i.e., converges to a feasible solution) within 134 iterations. Fig. 4 depicts the convergent behaviors of $\phi_\lambda(X)$ the CNO-CC algorithm on the ten datasets, where the red envelopes depict the augmented objective functions of group-best solutions X^*. It shows that the augmented objective function values monotonically decrease, and CNO-CC converges within 600 iterations.

D. Performance Comparisons

To make fair comparisons, the same stop criterion is used among the competing algorithms; i.e., M in CNO-CC is set according to the stop criterion. Fig. 5 shows the resulting clusters using the CNO-CC algorithm on the ten datasets. Note that the shapes of some clusters are irregular due to the existence of capacity constraints. Tables III–V records the mean values and standard deviations of 21 internal cluster validity criteria resulting from the seven competing algorithms over 20 runs with random initialization on the ten datasets. The results show that the CNO-CC algorithm achieves the best results among the ten methods in 126 out of the 168 cases (75%) in terms of the mean values.

VI. Conclusion

This paper presents a capacitated clustering algorithm based on CNO. The proposed objective function with fractional functional terms empowers to measure cluster compactness naturally. The surrogate function used to represent the objective function leads to the iteratively reweighted quadratic unconstrained binary optimization problem formulation in a majorization-minimization framework, facilitating the subsequent development of the clustering algorithm. The proposed capacitated clustering algorithm leverages the hill-climbing local search capability of BMs in scattered searches. The proposed algorithm statistically outperforms the baselines owing to the combined use of a more reasonable objective function for measuring cluster compactness and a more effective optimizer driven by collaborative neurodynamics. Further investigations may aim at the efficiency and scalability improvements of the neurodynamics-driven capacitated clustering algorithm. Further investigations may also include the robustness analysis of neurodynamics-driven constrained clustering as in noise-tolerant neural networks [112], [113].

References

[1] A. Banerjee and J. Ghosh, "Frequency-sensitive competitive learning for scalable balanced clustering on high-dimensional hyperspheres," *IEEE Trans. Neural Netw.*, vol. 15, no. 3, pp. 702–719, May 2004.

[2] H. Liu, J. Han, F. Nie, and X. Li, "Balanced clustering with least square regression," in *Proc. 31st AAAI Conf. Artif. Intell.*, 2017, pp. 2231–2237.

[3] J. Han, H. Liu, and F. Nie, "A local and global discriminative framework and optimization for balanced clustering," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 10, pp. 3059–3071, Oct. 2019.

[4] Z. Li, F. Nie, X. Chang, Z. Ma, and Y. Yang, "Balanced clustering via exclusive lasso: A pragmatic approach," in *Proc. AAAI Conf. Artif. Intell.*, vol. 32, no. 1, 2018, pp. 3596–3601.

[5] S. Ding, L. Cong, Q. Hu, H. Jia, and Z. Shi, "A multiway p-spectral clustering algorithm," *Knowl-Based Syst.*, vol. 164, pp. 371–377, Jan. 2019.

[6] P. Zhou, J. Chen, M. Fan, L. Du, Y.-D. Shen, and X. Li, "Unsupervised feature selection for balanced clustering," *Knowl.-Based Syst.*, vol. 193, Apr. 2020, Art. no. 105417.

[7] P. Li, H. Zhao, and H. Liu, "Deep fair clustering for visual learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 9070–9079.

[8] L. E. Caraballo, J.-M. Díaz-Báñez, and N. Kroher, "A polynomial algorithm for balanced clustering via graph partitioning," *Eur. J. Oper. Res.*, vol. 289, no. 2, pp. 456–469, Mar. 2021.

[9] J. M. Mulvey and M. P. Beck, "Solving capacitated clustering problems," *Eur. J. Oper. Res.*, vol. 18, pp. 339–348, Dec. 1984.

[10] V. Maniezzo, A. Mingozzi, and R. Baldacci, "A bionomic approach to the capacitated p-median problem," *J. Heuristics*, vol. 4, no. 3, pp. 263–280, 1998.

[11] S. Ahmadi and I. H. Osman, "Greedy random adaptive memory programming search for the capacitated clustering problem," *Eur. J. Oper. Res.*, vol. 162, no. 1, pp. 30–44, 2005.

[12] S. Scheuerer and R. Wendolsky, "A scatter search heuristic for the capacitated clustering problem," *Eur. J. Oper. Res.*, vol. 169, no. 2, pp. 533–547, Mar. 2006.

[13] K. Fleszar and K. S. Hindi, "An effective VNS for the capacitated p-median problem," *Eur. J. Oper. Res.*, vol. 191, no. 3, pp. 612–622, Dec. 2008.

[14] D. Rodney, A. J. Soper, and C. Walshaw, "Multilevel approaches applied to the capacitated clustering problem," in *Proc. Int. Conf. Sci. Comput.*, 2008, pp. 271–277.

[15] A. A. Chaves and L. A. N. Lorena, "Clustering search algorithm for the capacitated centered clustering problem," *Comput. Oper. Res.*, vol. 37, no. 3, pp. 552–558, Mar. 2010.

[16] Y. Deng and J. F. Bard, "A reactive GRASP with path relinking for capacitated clustering," *J. Heuristics*, vol. 17, no. 2, pp. 119–152, Apr. 2011.

[17] Z. Yang, H. Chen, and F. Chu, "A Lagrangian relaxation approach for a large scale new variant of capacitated clustering problem," *Comput. Ind. Eng.*, vol. 61, no. 2, pp. 430–435, Sep. 2011.

[18] M. Gallego, M. Laguna, R. Martí, and A. Duarte, "Tabu search with strategic oscillation for the maximally diverse grouping problem," *J. Oper. Res. Soc.*, vol. 64, no. 5, pp. 724–734, 2013.

[19] M. Lewis, H. Wang, and G. Kochenberger, "Exact solutions to the capacitated clustering problem: A comparison of two models," *Ann. Data Sci.*, vol. 1, no. 1, pp. 15–23, Mar. 2014.

[20] A. Martínez-Gavara, V. Campos, M. Gallego, M. Laguna, and R. Martí, "Tabu search and GRASP for the capacitated clustering problem," *Comput. Optim. Appl.*, vol. 62, no. 2, pp. 589–607, Nov. 2015.

[21] X. Lai and J.-K. Hao, "Iterated variable neighborhood search for the capacitated clustering problem," *Eng. Appl. Artif. Intell.*, vol. 56, pp. 102–120, Nov. 2016.

[22] A. A. Chaves, J. F. Gonçalves, and L. A. N. Lorena, "Adaptive biased random-key genetic algorithm with local search for the capacitated centered clustering problem," *Comput. Ind. Eng.*, vol. 124, pp. 331–346, Oct. 2018.

[23] F. Mai, M. J. Fry, and J. W. Ohlmann, "Model-based capacitated clustering with posterior regularization," *Eur. J. Oper. Res.*, vol. 271, no. 2, pp. 594–605, Dec. 2018.

[24] Q. Zhou, U. Benlic, Q. Wu, and J.-K. Hao, "Heuristic search to the capacitated clustering problem," *Eur. J. Oper. Res.*, vol. 273, no. 2, pp. 464–487, Mar. 2019.

[25] J. Brimberg, N. Mladenović, R. Todosijević, and D. Urošević, "Solving the capacitated clustering problem with variable neighborhood search," *Ann. Oper. Res.*, vol. 272, nos. 1–2, pp. 289–321, Jan. 2019.

[26] X. Lai, J.-K. Hao, Z.-H. Fu, and D. Yue, "Neighborhood decomposition-driven variable neighborhood search for capacitated clustering," *Comput. Oper. Res.*, vol. 134, May 2021, Art. no. 105362.

[27] K. Wagstaff and C. Cardie, "Clustering with instance-level constraints," in *Proc. AAAI/IAAI*, vol. 1097, Jun. 2000, pp. 577–584.

[28] Y. Jia, J. Hou, and S. Kwong, "Constrained clustering with dissimilarity propagation-guided graph-Laplacian PCA," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 9, pp. 3985–3997, Sep. 2021.

[29] Z. Li, F. Nie, X. Chang, L. Nie, H. Zhang, and Y. Yang, "Rank-constrained spectral clustering with flexible embedding," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 12, pp. 6073–6082, Dec. 2018.

[30] Y. A. Koskosidis and W. B. Powell, "Clustering algorithms for consolidation of customer orders into vehicle shipments," *Transp. Res. B, Methodol.*, vol. 26, no. 5, pp. 365–379, Oct. 1992.

[31] H. Ewbank, P. Wanke, and A. Hadi-Vencheh, "An unsupervised fuzzy clustering approach to the capacitated vehicle routing problem," *Neural Comput. Appl.*, vol. 27, no. 4, pp. 857–867, May 2016.

[32] F. Alesiani, G. Ermis, and K. Gkiotsalitis, "Constrained clustering for the capacitated vehicle routing problem (CC-CVRP)," *Appl. Artif. Intell.*, vol. 36, no. 1, pp. 1–25, Dec. 2022.

[33] Z. Donovan, K. Subramani, and V. Mkrtchyan, "Analyzing clustering and partitioning problems in selected VLSI models," *Theory Comput. Syst.*, vol. 64, no. 7, pp. 1242–1272, Oct. 2020.

[34] J. F. Bard and A. I. Jarrah, "Large-scale constrained clustering for rationalizing pickup and delivery operations," *Transp. Res. B, Methodol.*, vol. 43, no. 5, pp. 542–561, Jun. 2009.

[35] C.-A. Chou, W. A. Chaovalitwongse, T. Y. Berger-Wolf, B. DasGupta, and M. V. Ashley, "Capacitated clustering problem in computational biology: Combinatorial and statistical approach for sibling reconstruction," *Comput. Oper. Res.*, vol. 39, no. 3, pp. 609–619, Mar. 2012.

[36] A. Martínez-Gavara, D. Landa-Silva, V. Campos, and R. Martí, "Randomized heuristics for the capacitated clustering problem," *Inf. Sci.*, vol. 417, pp. 154–168, Nov. 2017.

[37] K. Liao and D. Guo, "A clustering-based approach to the capacitated facility location problem," *Trans. GIS*, vol. 12, no. 3, pp. 323–339, Jun. 2008.

[38] M. J. Negreiros, N. Maculan, P. L. Batista, J. A. Rodrigues, and A. W. Palhano, "Capacitated clustering problems applied to the layout of IT-teams in software factories," *Ann. Oper. Res.*, vol. 316, pp. 1157–1185, Sep. 2020.

[39] Y. Chen, Z.-P. Fan, J. Ma, and S. Zeng, "A hybrid grouping genetic algorithm for reviewer group construction problem," *Expert Syst. Appl.*, vol. 38, no. 3, pp. 2401–2411, Mar. 2011.

[40] Y. Bejerano, "Efficient integration of multihop wireless and wired networks with QoS constraints," *IEEE/ACM Trans. Netw.*, vol. 12, no. 6, pp. 1064–1078, Dec. 2004.

[41] R. Zhang, H. Li, and J. Wang, "Index tracking based on dynamic time warping and constrained k-medoids clustering," in *Proc. 11th Int. Conf. Intell. Control Inf. Process. (ICICIP)*, Dec. 2021, pp. 352–359.

[42] M. R. Garey and D. S. Johnson, *Comput. Intractability: A Guide to Theory NP-Completeness*, vol. 174. San Francisco, CA, USA: Freeman, 1979.

[43] J. J. Hopfield and D. W. Tank, "Computing with neural circuits—A model," *Science*, vol. 233, no. 4764, pp. 625–633, 1986.

[44] Z. Yan, J. Wang, and G. Li, "A collective neurodynamic optimization approach to bound-constrained nonconvex optimization," *Neural Netw.*, vol. 55, pp. 20–29, Jul. 2014.

[45] Z. Yan, J. Fan, and J. Wang, "A collective neurodynamic approach to constrained global optimization," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 5, pp. 1206–1215, May 2017.

[46] H. Che and J. Wang, "A collaborative neurodynamic approach to global and combinatorial optimization," *Neural Netw.*, vol. 114, pp. 15–27, Jun. 2019.

[47] H. Che and J. Wang, "A two-timescale duplex neurodynamic approach to biconvex optimization," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 8, pp. 2503–2514, Aug. 2019.

[48] H. Che and J. Wang, "A two-timescale duplex neurodynamic approach to mixed-integer optimization," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 1, pp. 36–48, Jan. 2021.

[49] J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," *Proc. Nat. Acad. Sci. USA*, vol. 79, no. 8, pp. 2554–2558, 1982.

[50] Y. Takefuji and K. C. Lee, "A near-optimum parallel planarization algorithm," *Science*, vol. 245, no. 4923, pp. 1221–1223, 1989.

[51] Y. Takefuji and K. C. Lee, "Artificial neural networks for four-coloring map problems and K-colorability problems," *IEEE Trans. Circuits Syst.*, vol. 38, no. 3, pp. 326–333, Mar. 1991.

[52] G. Galan-Marin and J. Munoz-Perez, "Design and analysis of maximum Hopfield networks," *IEEE Trans. Neural Netw.*, vol. 12, no. 2, pp. 329–339, Mar. 2001.

[53] Y. Takefuji and L. Chen, "Parallel algorithms for finding a near-maximum independent set of a circle graph," *IEEE Trans. Neural Netw.*, vol. 1, no. 3, p. 263, 1990.

[54] Y. Takefuji and K.-C. Lee, "A parallel algorithm for tiling problems," *IEEE Trans. Neural Netw.*, vol. 1, no. 1, pp. 143–145, Mar. 1990.

[55] Y. Takefuji and K.-C. Lee, "A super-parallel sorting algorithm based on neural networks," *IEEE Trans. Circuits Syst.*, vol. 37, no. 11, pp. 1425–1429, Nov. 1990.

[56] Y. Takefuji, T. Tanaka, and K. C. Lee, "A parallel string search algorithm," *IEEE Trans. Syst., Man, Cybern.*, vol. 22, no. 2, pp. 332–336, Mar. 1992.

[57] N. Funabikiy and Y. Takefuji, "A neural network parallel algorithm for channel assignment problems in cellular radio networks," *IEEE Trans. Veh. Technol.*, vol. 41, no. 4, pp. 430–437, Nov. 1992.

[58] S. C. Amartur, D. Piraino, and Y. Takefuji, "Optimization neural networks for the segmentation of magnetic resonance images," *IEEE Trans. Med. Imag.*, vol. 11, no. 2, pp. 215–220, Jun. 1992.

[59] K. C. Lee, N. Funabikiy, and Y. Takefuji, "A parallel improvement algorithm for the bipartite subgraph problem," *IEEE Trans. Neural Netw.*, vol. 3, no. 1, pp. 139–145, Jan. 1992.

[60] N. Funabikiy and Y. Takefuji, "A parallel algorithm for broadcast scheduling problems in packet radio networks," *IEEE Trans. Commun.*, vol. 41, no. 6, pp. 828–831, Jun. 1993.

[61] K. Tsuchiya, S. Bharitkar, and Y. Takefuji, "A neural network approach to facility layout problems," *Eur. J. Oper. Res.*, vol. 89, no. 3, pp. 556–563, Mar. 1996.

[62] N. Funabiki and S. Nishikawa, "A neural network model for multilayer topological via minimization in a switchbox," *IEEE Trans. Comput.-Aided Design Integr.*, vol. 15, no. 8, pp. 1012–1020, Aug. 1996.

[63] K. Tsuchiya and Y. Takefuji, "A neural network approach to PLA folding problems," *IEEE Trans. Comput.-Aided Design Integr.*, vol. 15, no. 10, pp. 1299–1305, Oct. 1996.

[64] S. Bharitkar, K. Tsuchiya, and Y. Takefuji, "Microcode optimization with neural networks," *IEEE Trans. Neural Netw.*, vol. 10, no. 3, pp. 698–703, May 1999.

[65] G. E. Hinton and T. J. Sejnowski, "Optimal perceptual inference," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 1983, pp. 448–453.

[66] E. H. L. Aarts and J. H. M. Korst, "Boltzmann machines as a model for parallel annealing," *Algorithmica*, vol. 6, nos. 1–6, pp. 437–465, Jun. 1991.

[67] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, no. 4598, pp. 671–680, 1983.

[68] J. H. M. Korst and E. H. L. Aarts, "Combinatorial optimization on a Boltzmann machine," *J. Parallel Distrib. Comput.*, vol. 6, no. 2, pp. 331–357, Apr. 1989.

[69] E. H. L. Aarts and J. H. M. Korst, "Boltzmann machines for travelling salesman problems," *Eur. J. Oper. Res.*, vol. 39, no. 1, pp. 79–95, Mar. 1989.

[70] Q. Liu, S. Yang, and J. Wang, "A collective neurodynamic approach to distributed constrained optimization," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 8, pp. 1747–1758, Aug. 2017.

[71] S. Yang, Q. Liu, and J. Wang, "A collaborative neurodynamic approach to multiple-objective distributed optimization," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 4, pp. 981–992, Apr. 2018.

[72] M.-F. Leung and J. Wang, "A collaborative neurodynamic approach to multiobjective optimization," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 11, pp. 5738–5748, Nov. 2018.

[73] H. Che, J. Wang, and A. Cichocki, "Bicriteria sparse nonnegative matrix factorization via two-timescale duplex neurodynamic optimization," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Nov. 17, 2021, doi: 10.1109/TNNLS.2021.3125457.

[74] J. Wang, J. Wang, and H. Che, "Task assignment for multivehicle systems based on collaborative neurodynamic optimization," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 4, pp. 1145–1154, Apr. 2020.

[75] J. Wang, J. Wang, and Q.-L. Han, "Multivehicle task assignment based on collaborative neurodynamic optimization with discrete Hopfield networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 12, pp. 5274–5286, Dec. 2021.

[76] M.-F. Leung and J. Wang, "Minimax and biobjective portfolio selection based on collaborative neurodynamic optimization," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 7, pp. 2825–2836, Jul. 2021.

[77] J. Zhao, J. Yang, J. Wang, and W. Wu, "Spiking neural network regularization with fixed and adaptive drop-keep probabilities," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 8, pp. 4096–4109, Aug. 2022.

[78] X. Li, J. Wang, and S. Kwong, "Hash bit selection via collaborative neurodynamic optimization with discrete Hopfield networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 10, pp. 5116–5124, 2022.

[79] W. Zhou, H.-T. Zhang, and J. Wang, "Sparse Bayesian learning based on collaborative neurodynamic optimization," *IEEE Trans. Cybern.*, early access, Jul. 14, 2021, doi: 10.1109/TCYB.2021.3090204.

[80] J. Kennedy and R. C. Eberhart, "A discrete binary version of the particle swarm algorithm," in *Proc. IEEE Int. Conf. Syst., Man, Cybern. Comput. Simulation*, vol. 5, Oct. 1997, pp. 4104–4108.

[81] D. R. Hunter and K. Lange, "A tutorial on MM algorithms," *Amer. Statist.*, vol. 58, no. 1, pp. 30–37, 2004.

[82] C. Lu, J. Feng, S. Yan, and Z. Lin, "A unified alternating direction method of multipliers by majorization minimization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 3, pp. 527–541, Mar. 2018.

[83] Y. Sun, P. Babu, and D. Palomar, "Majorization-minimization algorithms in signal processing, communications, and machine learning," *IEEE Trans. Signal Process.*, vol. 65, no. 3, pp. 794–816, Feb. 2016.

[84] Z. Lin, C. Xu, and H. Zha, "Robust matrix factorization by majorization minimization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 1, pp. 208–220, Jan. 2017.

[85] F. Nie, Z. Wang, L. Tian, R. Wang, and X. Li, "Subspace sparse discriminative feature selection," *IEEE Trans. Cybern.*, vol. 52, no. 6, pp. 4221–4233, Jun. 2022.

[86] G. Kochenberger et al., "The unconstrained binary quadratic programming problem: A survey," *J. Combinat. Optim.*, vol. 28, no. 1, pp. 58–81, Jul. 2014.

[87] M. Negreiros and A. Palhano, "The capacitated centred clustering problem," *Comput. Oper. Res.*, vol. 33, no. 6, pp. 1639–1663, Jun. 2006.

[88] A. A. Chaves and L. A. N. Lorena, "Hybrid evolutionary algorithm for the capacitated centered clustering problem," *Expert Syst. Appl.*, vol. 38, no. 5, pp. 5013–5018, May 2011.

[89] F. Stefanello, O. C. B. de Arújo, and F. M. Müller, "Matheuristics for the capacitated p-median problem," *Int. Trans. Oper. Res.*, vol. 22, pp. 149–167, Jan. 2015.

[90] S.-O. Caballero-Morales, E. Barojas-Payan, D. Sanchez-Partida, and J.-L. Martinez-Flores, "Extended GRASP-capacitated K-means clustering algorithm to establish humanitarian support centers in large regions at risk in Mexico," *J. Optim.*, vol. 2018, pp. 1–14, Dec. 2018.

[91] A. E. F. Muritiba, M. J. N. Gomes, M. F. de Souza, and H. L. G. Oriá, "Path-relinking with Tabu search for the capacitated centered clustering problem," *Expert Syst. Appl.*, vol. 198, Jul. 2022, Art. no. 116766.

[92] Y. Xu, P. Guo, and Y. Zeng, "An iterative neighborhood local search algorithm for capacitated centered clustering problem," *IEEE Access*, vol. 10, pp. 34497–34510, 2022.

[93] J. O. McClain and V. R. Rao, "CLUSTISz: A program to test for the quality of clustering of a set of objects," *J. Marketing Res.*, vol. 12, no. 4, pp. 456–460, 1975.

[94] F. J. Rohlf, "Methods of comparing classifications," *Annu. Rev. Ecology Systematics*, vol. 5, no. 1, pp. 101–113, Nov. 1974.

[95] F. B. Baker and L. J. Hubert, "Measuring the power of hierarchical cluster analysis," *J. Amer. Stat. Assoc.*, vol. 70, no. 349, pp. 31–38, Mar. 1975.

[96] L. Hubert and J. Schultz, "Quadratic assignment as a general data analysis strategy," *Brit. J. Math. Stat. Psychol.*, vol. 29, no. 2, pp. 190–241, Nov. 1976.

[97] B. Desgraupes, "Clustering indices," Univ. Paris Ouest-Lab Modal'X, Nanterre, France, Tech. Rep., 2013, p. 34, vol. 1.

[98] J. C. Bezdek and N. R. Pal, "Some new indexes of cluster validity," *IEEE Trans. Syst., Man, Cybern. B. Cybern.*, vol. 28, no. 3, pp. 301–315, Jun. 1998.

[99] D. Ratkowsky and G. Lance, "Criterion for determining the number of groups in a classification," *Austral. Comput. J.*, vol. 10, no. 3, pp. 115–117, 1978.

[100] T. Caliński and J. Harabasz, "A dendrite method for cluster analysis," *Commun. Stat., Theory Methods*, vol. 3, no. 1, pp. 1–27, 1974.

[101] S. Ray and R. H. Turi, "Determination of number of clusters in k-means clustering and application in colour image segmentation," in *Proc. 4th Int. Conf. Adv. Pattern Recognit. Digit. Techn.*, 1999, pp. 137–143.

[102] J. C. Dunn, "Well-separated clusters and optimal fuzzy partitions," *J. Cybern.*, vol. 4, no. 1, pp. 95–104, 2008.

[103] H. P. Friedman and J. Rubin, "On some invariant criteria for grouping data," *J. Amer. Statist. Assoc.*, vol. 62, no. 320, pp. 1159–1178, 1967.

[104] G. Ball and D. Hall, *A Novel Method of data analysis and Pattern Classification*. Menlo Park, CA, USA: SRI International, 1965.

[105] G. W. Milligan, "A Monte Carlo study of thirty internal criterion measures for cluster analysis," *Psychometrika*, vol. 46, no. 2, pp. 187–199, Jun. 1981.

[106] X. L. Xie and G. Beni, "A validity measure for fuzzy clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 8, pp. 841–847, Aug. 1991.

[107] D. L. Davies and D. W. Bouldin, "A cluster separation measure," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-1, no. 2, pp. 224–227, Apr. 1979.

[108] A. J. Scott and M. J. Symons, "Clustering methods based on likelihood ratio criteria," *Biometrics*, vol. 27, no. 2, pp. 387–397, 1971.

[109] F. H. C. Marriott, "Practical problems in a method of cluster analysis," *Biometrics*, vol. 27, no. 3, pp. 501–514, 1971.

[110] J. A. Hartigan, *Clustering Algorithms*. Hoboken, NJ, USA: Wiley, 1975.

[111] A. W. Edwards and L. L. Cavalli-Sforza, "A method for cluster analysis," *Biometrics*, vol. 21, no. 2, pp. 362–375, 1965.

[112] L. Xiao, J. Dai, R. Lu, S. Li, J. Li, and S. Wang, "Design and comprehensive analysis of a noise-tolerant ZNN model with limited-time convergence for time-dependent nonlinear minimization," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 12, pp. 5339–5348, Dec. 2020.

[113] L. Xiao, Y. He, J. Dai, X. Liu, B. Liao, and H. Tan, "A variable-parameter noise-tolerant zeroing neural network for time-variant matrix inversion with guaranteed robustness," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 4, pp. 1535–1545, Apr. 2022.

Hongzong Li received the B.E. degree in automation from Northeastern University, Shenyang, Liaoning, China, in 2020. He is currently pursuing the Ph.D. degree in computer science with the Department of Computer Science, City University of Hong Kong, Hong Kong.

His current research interests include optimization, computational intelligence, and clustering.

Jun Wang (Life Fellow, IEEE) received the B.S. and M.S. degrees from Dalian University of Technology, Dalian, China, in 1982 and 1985, respectively, and the Ph.D. degree from Case Western Reserve University, Cleveland, OH, USA, in 1991.

He held various academic positions at Dalian University of Technology, Case Western Reserve University, University of North Dakota, Grand Forks, ND, USA, and The Chinese University of Hong Kong, Hong Kong. He also held various short-term or part-time visiting positions at the U.S. Air Force Armstrong Laboratory, Dayton, OH, USA, RIKEN Brain Science Institute, Tokyo, Japan, the Huazhong University of Science and Technology, Wuhan, China, Shanghai Jiao Tong University, Shanghai, China, and Dalian University of Technology. He is currently a Chair Professor with the City University of Hong Kong, Hong Kong.

Dr. Wang was a Distinguished Lecturer of the IEEE Computational Intelligence Society from 2010 to 2012 and from 2014 to 2016. He is an IEEE Systems, Man, and Cybernetics Society Distinguished Lecturer from 2017 to 2022. He was a recipient of several awards, such as the Research Excellence Award from the Chinese University of Hong Kong for 2008–2009, the Outstanding Achievement Award from the Asia–Pacific Neural Network Assembly in 2011, the IEEE TRANSACTIONS ON NEURAL NETWORKS Outstanding Paper Award in 2011, the Neural Networks Pioneer Award from the IEEE Computational Intelligence Society in 2014, and the Norbert Wiener Award from the IEEE Systems, Man, and Cybernetics Society in 2019. He served as the General Chair for the 13th and 25th International Conference on Neural Information Processing in 2006 and 2018, respectively, and the IEEE World Congress on Computational Intelligence in 2008. He was the Editor-in-Chief of the IEEE TRANSACTIONS ON CYBERNETICS from 2014 to 2019.