Knowledge-Based Systems 255 (2022) 109593

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

Collaborative annealing power *k*-means++ clustering☆

Hongzong Li [a], Jun Wang [a,b,*]

[a] Department of Computer Science, City University of Hong Kong, Kowloon, Hong Kong
[b] School of Data Science, City University of Hong Kong, Kowloon, Hong Kong

ARTICLE INFO

Article history:
Received 31 May 2022
Received in revised form 5 July 2022
Accepted 30 July 2022
Available online 24 August 2022

Keywords:
k-means clustering
k-means++
Power *k*-means
Collaborative neurodynamic optimization

ABSTRACT

Clustering is the most fundamental technique for data processing. This paper presents a collaborative annealing power *k*-means++ clustering algorithm by integrating the *k*-means++ and power *k*-means algorithms in a collaborative neurodynamic optimization framework. The proposed algorithm starts with *k*-means++ to select initial cluster centers, then leverages the power *k*-means to find multiple sets of centers as alternatives and a particle swarm optimization rule to reinitialize the centers in the subsequential iterations for improving clustering performance. Experimental results on twelve benchmark datasets are elaborated to demonstrate the superior performance of the proposed algorithm to seven mainstream clustering algorithms in terms of 21 internal and external indices.

1. Introduction

Clustering plays an important role in data analysis and processing. It aims to group unlabeled data into a number of clusters such that the data within each cluster are similar to each other [1]. It has been widely applied to many areas, such as data mining, image segmentation, community detection, homology identification, and market analysis, among many others [2].

Existing clustering algorithms may be classified into multiple categories from several perspectives. According to the structure of algorithms, they may be classified into two categories [3]: hierarchical clustering and partitional clustering algorithms. Hierarchical clustering methods are further subdivided into agglomerative algorithms [4] and divisive algorithms [5]. Partitional clustering methods are further subdivided into center-based algorithms and center-less algorithms. Center-based algorithms are exemplified by *k*-means [6], *k*-harmonic means [7], *k*-medoids algorithms [8], and spectral clustering algorithms [9–11]. The center-less algorithms include distribution-based and densities-based algorithms, etc. The distribution-based methods include the expectation–maximization for Gaussian mixture model algorithms [12], and the density peak with Gaussian mixture model algorithm [13], etc. The density-based methods include the mean-shift algorithm [14], the density peak algorithm [15], etc. Depending on the uniqueness of sample-cluster assignments, clustering algorithms may be classified into hard and soft clustering algorithms. The soft clustering methods include probabilistic clustering algorithms [16], fuzzy clustering algorithms [17–20]. In terms of feature space, clustering may operate in whole space, weighted space, or subspace. Feature-weighted clustering methods include the entropy weighting *k*-means clustering algorithm [21], the entropy-weighted power *k*-means clustering algorithm [22], the lasso-weighted *k*-means clustering algorithm [23]. Subspace clustering methods include the neural collaborative subspace clustering algorithm [24], the multi-view subspace clustering algorithm [25], and the subspace clustering via adaptive least square regression with smooth affinities algorithm [26]. In addition, many computational intelligence methods are proposed for clusterings based on tabu search [27], simulated annealing [28], collaborative neurodynamic optimization [29], etc.

Of particular interest, the classic *k*-means algorithm is commonly used and is still one of the most widely used methods nowadays due to its efficiency and simplicity. However, a popular implementation of *k*-means, Lloyd's algorithm [6] is sensitive to initial centers and it deteriorates in high dimensions [47,48]. To overcome these drawbacks, many improvements have been proposed, such as initialization improvements [8,49–51] and objective function improvements [7,48,52–54]. In particular, the power *k*-means clustering algorithm [48] operates by minimizing a power-mean function with annealing parameters to mitigate its tendency to get trapped in local minima. Nevertheless, its clustering performance still depends on the initial centers.

In this paper, we propose a clustering algorithm (called collaborative annealing power *k*-means++ clustering) that integrates

☆ This work was supported in part by the Research Grants Council of the Hong Kong Special Administrative Region of China under Grants 11202318, 11202019, and 11203721; and in part by the InnoHK initiative, Hong Kong, the Government of the Hong Kong Special Administrative Region, Hong Kong, and Laboratory for AI-Powered Financial Technologies, Hong Kong.

* Corresponding author at: Department of Computer Science, City University of Hong Kong, Kowloon, Hong Kong.
E-mail addresses: hongzli2-c@my.cityu.edu.hk (H. Li), jwang.cs@cityu.edu.hk (J. Wang).

https://doi.org/10.1016/j.knosys.2022.109593
0950-7051/

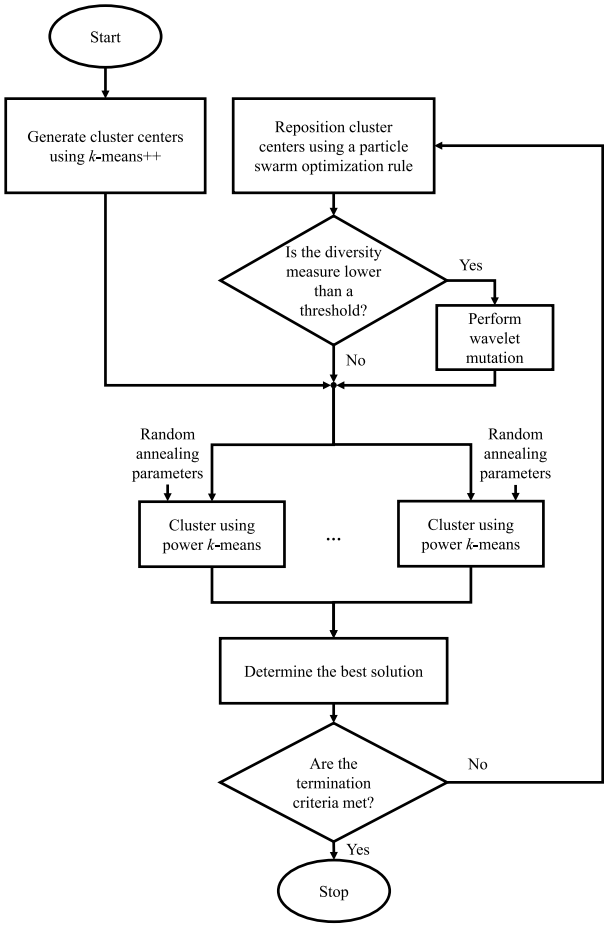

Fig. 1. A flow chart of collaborative annealing power k-means++ clustering.

Table 1
Information about the twelve benchmark datasets and the corresponding hyper-parameters value in CAPKM++ used in the experiments.

Dataset	Type	n	m	k	M	N
NCI9[4.2]	biological data	60	9712	9	30	40
Lymphoma[4.2]	biological data	96	4026	9	30	10
ORL10P[4.2]	face image data	100	10304	10	30	4
WarpPIE10P[4.2]	face image data	210	2420	10	30	10
Segment[4.2]	image segmentation data	2310	19	7	30	3
SpamBase[4.2]	e-mail message data	4597	57	2	5	5
PageBlocks[4.2]	text data	5472	10	5	30	12
Texture[4.2]	texture data	5500	40	11	15	40
Optdigits[4.2]	hand-written image data	5620	65	10	30	40
Satimage[4.2]	satellite image data	6435	36	7	30	5
COIL2000[4.2]	customer information data	9822	85	2	5	2
Penbased[4.2]	hand-written image data	10992	16	10	15	2

the k-mean++ and power k-means algorithms in a collaborative neurodynamic optimization framework. In the beginning, k-means++ is used to provide good initial centers. Power k-means with randomly generated annealing parameters is leveraged to generate multiple centers for alternative clustering, and a particle swarm optimization rule is used for repositioning the centers. Attributable to the joint effects of the better seeding from k-means++, the smoother objective function from Power k-means, the more diversified alternatives associated with the multiple centers repositioned using the swarm intelligence method, the proposed clustering algorithm could produce better clustering results than existing k-means algorithms.

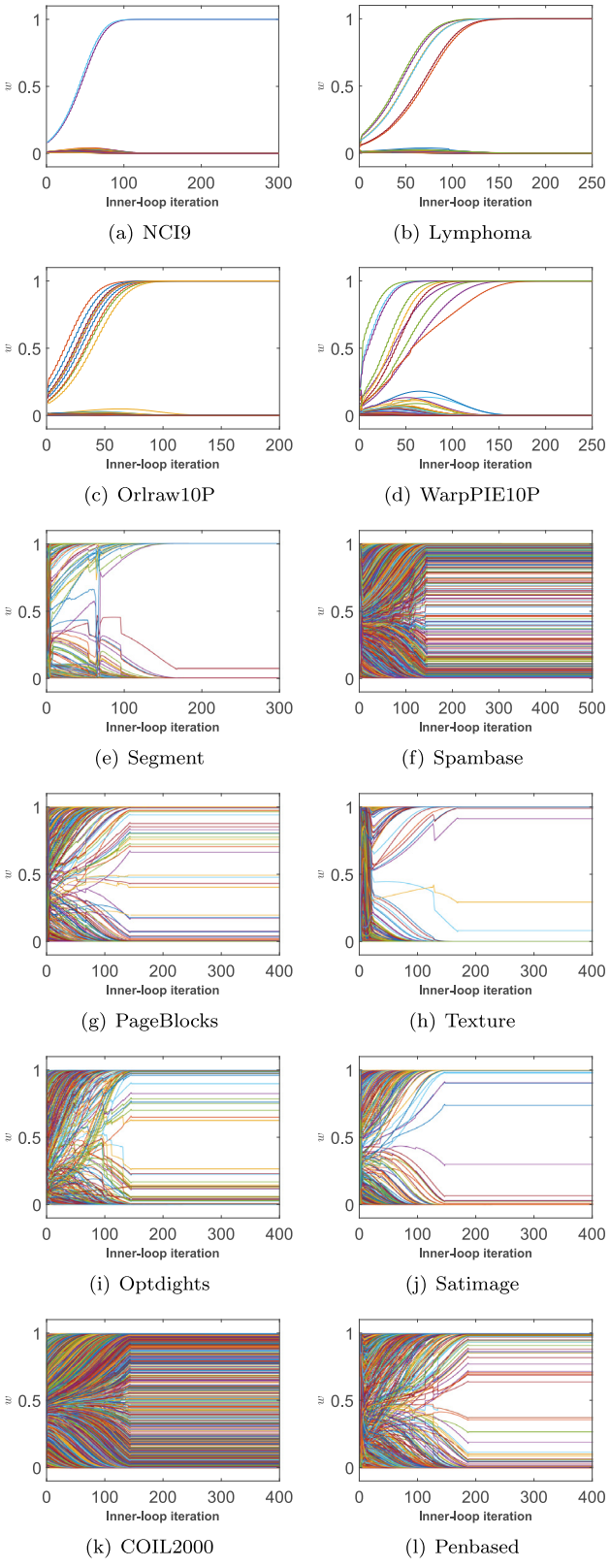

Fig. 2. Snapshots of the weight vectors w in (8) in the inner loop (Steps 4–8) of CAPKM++ on the twelve datasets.

The rest of this paper is organized as follows. In Section 2, the problem statement, k-means++ and power k-means clustering algorithms, and collaborative neurodynamic optimization are

Table 2
The mean values and standard deviations of $f(\Theta)$ by using PKM, PKM++, CAPKM, and CAPKM++ in the ablation studies.

Dataset	PKM	PKM++	CAPKM	CAPKM++
NCI9	4.8597 ± 0.0106	4.8580 ± 0.0095	4.8362 ± 0.0081	**4.8105 ± 0**
Lymphoma	9.4485 ± 0.0088	9.4469 ± 0.0070	9.4397 ± 0.0017	**9.4388 ± 0**
ORL10P	2.4151 ± 0	2.4130 ± 0.0061	**2.3809 ± 0**	**2.3809 ± 0**
WarpPIE10P	3.4688 ± 0.0093	3.4648 ± 0	3.4489 ± 0.0013	**3.4481 ± 0**
Segment	22.0185 ± 0.5957	22.2194 ± 0.6081	**21.4986 ± 0**	**21.4986 ± 0**
SpamBase	13.6374 ± 0	13.6374 ± 0	**13.4122 ± 0**	**13.4122 ± 0**
PageBlocks	21.6410 ± 0.0478	21.6002 ± 0.0594	**21.5478 ± 0**	**21.5478 ± 0**
Texture	15.6271 ± 0.0031	15.6107 ± 0.0200	15.5734 ± 0.0001	**15.5733 ± 0**
Optdigits	235.6059 ± 1.1878	235.0509 ± 0.1682	**234.8252** ± 0	**234.8252 ± 0**
Satimage	44.4344 ± 0.8616	44.2245 ± 0.6462	**44.0144 ± 0**	**44.0144 ± 0**
COIL2000	222.9554 ± 0	222.9554 ± 0	**222.9522 ± 0**	**222.9522 ± 0**
Penbased	317.3646 ± 4.1250	319.1628 ± 3.8478	308.1348 ± 0.0007	**308.1344 ± 0**

Table 3
Internal cluster validity indices.

Measure	Definition	Min/Max	Reference
WGSS	$\sum_{k=1}^{p}\sum_{x\in C_k}\|\|x-c_k\|\|^2$	min	[30]
Trace W Index (TWI)	$Tr(WG)$	max	[31]
Ball–Hall Index (BHI)	$\frac{1}{p}\sum_{k=1}^{p}\frac{1}{n_k}\sum_{x\in C_k}\|\|x-c_k\|\|^2$	max	[32]
G+ Index (GPI)	$\frac{2s^-}{N_T(N_T-1)}$	min	[33]
Calinski Harabasz Index (CHI)	$\frac{n-p}{p-1}\frac{BGSS}{WGSS}$	max	[34]
Dunn Index (DI)	$\frac{\min_{i,j}\{\min_{x\in C_i,y\in C_j} d(x,y)\}}{\max_k\{\max_{x,y\in C_k} d(x,y)\}}$	max	[35]
Baker–Hubert Gamma Index (BHGI)	$\frac{s^+-s^-}{s^++s^-}$	max	[36]
McClain–Rao Index (MRI)	$\frac{N_B}{N_W}\frac{S_W}{S_B}$	min	[37]
Log SS Ratio Index (LSSRI)	$\log(\frac{BGSS}{WGSS})$	min	[38]
C Index (CI)	$\frac{S_W-S_{\min}}{S_{\max}-S_{\min}}$	min	[39]
Ratkowsky–Lance Index (RLI)	$\sqrt{\frac{\sum_{j=1}^{m}(BGSS_j/TSS_j)/m}{p}}$	max	[40]
Davies Bouldin Index (DBI)	$\frac{1}{p}\sum_{i}^{n}\max_{j\neq i}\{\frac{\frac{1}{n_i}\sum_{x\in C_i}\|\|x-c_i\|\|+\frac{1}{n_j}\sum_{x\in C_j}\|\|x-c_j\|\|}{d(c_i,c_j)}\}$	min	[41]
PBM Index (PBMI)	$\left(\frac{1}{p}\times\frac{\sum_{x\in D}d(x,c)}{\sum_{k=1}^{p}\sum_{x\in C_k}d(x,c_k)}\times\max_{k<k'}d(c_k-c_{k'})\right)^2$	max	[42]
Xie–Beni Index (XBI)	$WGSS/[n\times\left(\min_{i<j}\{\min_{x\in C_i,y\in C_j}d(x,y)\}\right)^2]$	min	[43]
Dunn Generalized Index (DGI)	$\frac{\min_{k\neq k'}\delta(C_k,C_k')}{\max_k\Delta(C_k)}$	max	[44]
Ray–Turi Index (RTI)	$\frac{1}{n}\frac{WGSS}{\min_{k<k'}\|\|c_k-c_{k'}\|\|^2}$	min	[45]
Tau Index (TI)	$\frac{s^+-s^-}{\sqrt{N_BN_W(\frac{N_T(N_T-1)}{2})}}$	max	[46]
Wemmert–Gancarski Index (WGI)	$\frac{1}{n}\sum_{k=1}^{p}\max\{0,n_k-\sum_{x\in C_k}(\frac{\|\|x-c_k\|\|}{\min_{k\neq k'}\|\|x-c_{k'}\|\|})\}$	max	[46]

D: dataset; p: the number of clusters; c: center of D; C_i: the ith 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 $\frac{n(n-1)}{2}$; $S_{\min}$: the smallest within-cluster distance; $S_{\max}$: the largest within-cluster distance; $\Delta(C_k)$: the within-cluster distances $\frac{1}{n_k(n_k-1)}\sum_{x,y\in C_k,x\neq y}d(x,y)$; $\delta(C_k,C_k')$: 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: $\sum_{i=1}^{N}(D_{ij}-c^j)$; $BGSS$: $\sum_{k=1}^{p}n_k||c_k-c||^2$; $T:=t_{ij}=\sum_{x\in D}(x^i-c^i)(x^j-c^j)$; $WG^k\in\mathbb{R}^{p\times p}$ 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 $b_{ij}=\sum_{k=1}^{p}n_k(c_k^i-c_i)(c_k^j-c^j)$.

Table 4
The mean values and standard deviations of internal and external cluster validity indices resulting from CAPKM++ and seven baselines on NCI9 and Lymphoma.

NCI9	KM	KM++	PKM	PKM++	EWPKM	SC	HC	CAPKM++
WGSS↓	5.2345 ± 0.1907	5.2417 ± 0.1304	4.8597 ± 0.0106	4.858 ± 0.0095	4.8586 ± 0.0101	7.0303 ± 0.1475	5.0366 ± 0	**4.8105 ± 0**
MRI↓	0.8729 ± 0.0154	0.8699 ± 0.0117	0.8472 ± 0.0052	0.8459 ± 0.005	0.8457 ± 0.0054	1.0006 ± 0.0152	0.8442 ± 0	**0.8389 ± 0**
GPI↓	0.0476 ± 0.013	0.0465 ± 0.0116	0.0245 ± 0.0024	0.0242 ± 0.0022	0.0241 ± 0.0023	0.2402 ± 0.0207	0.0247 ± 0	**0.021 ± 0**
BHGI↑	0.638 ± 0.0747	0.6512 ± 0.0642	0.7386 ± 0.0338	0.7451 ± 0.0292	0.7452 ± 0.0324	0.0012 ± 0.0729	0.7685 ± 0	**0.7948 ± 0**
CI↓	0.1668 ± 0.0349	0.1587 ± 0.0264	0.1243 ± 0.0137	0.121 ± 0.0126	0.1208 ± 0.0138	0.4998 ± 0.0397	0.1102 ± 0	**0.1008 ± 0**
TI↑	0.3247 ± 0.0361	0.3339 ± 0.0328	0.321 ± 0.0207	0.3254 ± 0.0174	0.3246 ± 0.0194	0.0011 ± 0.0511	0.3551 ± 0	**0.3598 ± 0**
DGI↑	1.4282 ± 0.0813	1.4398 ± 0.0855	1.5146 ± 0.0224	1.5176 ± 0.022	1.5151 ± 0.0227	1.1752 ± 0.1038	**1.5442 ± 0**	1.5269 ± 0
RLI↑	0.1864 ± 0.0047	0.1878 ± 0.0056	0.1936 ± 0.0005	0.1937 ± 0.0004	0.1936 ± 0.0005	0.1214 ± 0.0051	0.1949 ± 0	**0.1964 ± 0**
CHI↑	0.4882 ± 0.0323	0.498 ± 0.0397	0.5548 ± 0.0034	0.5554 ± 0.003	0.5552 ± 0.0032	0.153 ± 0.0138	0.5531 ± 0	**0.5707 ± 0**
RTI↓	2.2602 ± 0.4071	2.0109 ± 0.337	1.8628 ± 0.0715	**1.8095 ± 0.093**	1.8153 ± 0.0908	5.8821 ± 3.3645	1.886 ± 0	1.8742 ± 0
WGI↑	0.189 ± 0.0138	0.2001 ± 0.0192	0.2085 ± 0.0039	0.2114 ± 0.0056	0.2099 ± 0.0047	0.136 ± 0.0236	0.2143 ± 0	**0.2249 ± 0**
DI↑	0.6168 ± 0.0288	0.619 ± 0.0378	0.6637 ± 0.0086	0.6617 ± 0.0102	0.6589 ± 0.0127	0.4688 ± 0.0105	**0.689 ± 0**	0.6779 ± 0
BHI↑	0.0761 ± 0.0038	0.0728 ± 0.0081	0.0801 ± 0.0004	0.08 ± 0.0008	**0.0802 ± 0.0005**	0.0506 ± 0.0107	0.0727 ± 0	0.0726 ± 0
PBMI↑	0.0049 ± 0.0012	0.0047 ± 0.0008	**0.0059 ± 0.0003**	0.0057 ± 0.0004	0.0058 ± 0.0003	0.005 ± 0.0006	0.0054 ± 0	0.005 ± 0
XBI↓	0.7165 ± 0.0673	0.6981 ± 0.0638	0.6136 ± 0.0013	0.6134 ± 0.0012	0.6135 ± 0.0013	0.9819 ± 0.0152	0.6143 ± 0	**0.6074 ± 0**
DBI↓	2.4802 ± 0.1045	2.3436 ± 0.2082	2.5172 ± 0.0524	2.4772 ± 0.0663	2.5064 ± 0.0561	2.4248 ± 0.4514	2.2793 ± 0	**2.2482 ± 0**
LSSRI↑	−0.7824 ± 0.08	−0.7568 ± 0.0783	−0.5891 ± 0.0061	−0.5881 ± 0.0055	−0.5885 ± 0.0058	−1.8809 ± 0.0892	−0.5921 ± 0	**−0.5608 ± 0**
TWI↓	0.5759 ± 0.0143	0.5713 ± 0.0142	0.54 ± 0.0012	0.5398 ± 0.0011	0.5398 ± 0.0011	0.7282 ± 0.0087	0.5406 ± 0	**0.5345 ± 0**
ACC↑	0.3867 ± 0.0345	0.4042 ± 0.0358	0.4317 ± 0.0152	0.4392 ± 0.0182	0.4317 ± 0.0131	0.2533 ± 0.0239	**0.4667 ± 0**	0.4333 ± 0
NMI↑	0.3967 ± 0.0438	0.4053 ± 0.0443	**0.4752 ± 0.0126**	0.4731 ± 0.015	0.4702 ± 0.0143	0.2488 ± 0.0265	0.475 ± 0	0.4658 ± 0
ARI↑	0.1049 ± 0.0387	0.1232 ± 0.0417	0.1767 ± 0.0135	0.1759 ± 0.0184	0.1706 ± 0.0158	−0.0017 ± 0.0117	**0.1946 ± 0**	0.1823 ± 0
Lymphoma	**KM**	**KM++**	**PKM**	**PKM++**	**EWPKM**	**SC**	**HC**	**CAPKM++**
WGSS↓	10.0905 ± 0.2695	10.0917 ± 0.2289	9.4485 ± 0.0088	9.4469 ± 0.0070	9.4598 ± 0.0257	13.6896 ± 0.1615	9.5129 ± 0	**9.4388 ± 0**
MRI↓	0.8614 ± 0.0160	0.8644 ± 0.0145	0.8356 ± 0.0009	0.8355 ± 0.0008	0.8356 ± 0.0016	1.0014 ± 0.0077	0.8426 ± 0	**0.8327 ± 0**
GPI↓	0.0339 ± 0.0150	0.0354 ± 0.0118	0.0171 ± 0.0005	0.0170 ± 0.0004	0.0170 ± 0.0010	0.2421 ± 0.0151	0.0212 ± 0	**0.0160 ± 0**
BHGI↑	0.7269 ± 0.0861	0.7197 ± 0.0772	0.8307 ± 0.0064	0.8309 ± 0.0019	0.831 ± 0.0110	−0.0104 ± 0.0471	0.7883 ± 0	**0.8358 ± 0**
CI↓	0.1635 ± 0.0411	0.1686 ± 0.0392	0.1125 ± 0.0032	0.1125 ± 0.0009	0.1127 ± 0.0050	0.4980 ± 0.0249	0.1320 ± 0	**0.1084 ± 0**
TI↑	0.3581 ± 0.0446	0.3602 ± 0.048	**0.3731 ± 0.0061**	0.3727 ± 0.0022	0.3731 ± 0.0074	−0.0073 ± 0.0326	0.3526 ± 0	0.3694 ± 0
DGI↑	1.3461 ± 0.1648	1.3328 ± 0.1485	1.5686 ± 0.0074	**1.5706 ± 0.003**	1.5360 ± 0.0885	1.0975 ± 0.0435	1.5685 ± 0	1.5694 ± 0
RLI↑	0.1811 ± 0.0058	0.1806 ± 0.0049	0.1939 ± 0.0002	0.1940 ± 0.0001	0.1937 ± 0.0005	0.0956 ± 0.0034	0.1925 ± 0	**0.1940 ± 0**
CHI↑	0.4267 ± 0.037	0.4233 ± 0.0321	0.5194 ± 0.0014	0.5197 ± 0.0011	0.5176 ± 0.0041	0.0897 ± 0.0071	0.5091 ± 0	**0.5210 ± 0**
RTI↓	2.6389 ± 0.3654	2.5379 ± 0.524	2.2287 ± 0.0815	2.216 ± 0.1035	2.2094 ± 0.1729	8.5753 ± 3.1523	2.0090 ± 0	**1.8269 ± 0**
WGI↑	0.1784 ± 0.0154	0.1762 ± 0.0199	0.2070 ± 0.0006	0.2072 ± 0.0005	0.2067 ± 0.0024	0.0818 ± 0.0122	0.2083 ± 0	**0.2107 ± 0**
DI↑	0.5729 ± 0.0733	0.5617 ± 0.0616	0.7007 ± 0.0082	**0.7013 ± 0.0088**	0.6834 ± 0.0435	0.4526 ± 0.0112	0.6745 ± 0	0.6745 ± 0
BHI↑	**0.0986 ± 0.0027**	0.0964 ± 0.004	0.0958 ± 0.0003	0.0959 ± 0.0002	0.0960 ± 0.0006	0.0682 ± 0.0114	0.0952 ± 0	0.0968 ± 0
PBMI↑	0.0036 ± 0.0004	0.0041 ± 0.0005	0.0042 ± 0.0001	0.0042 ± 0.0001	0.0042 ± 0.0001	**0.0049 ± 0.0007**	0.004 ± 0	0.0046 ± 0
XBI↓	0.9094 ± 0.2176	0.9332 ± 0.1899	0.6103 ± 0.0006	0.6102 ± 0.0005	0.6486 ± 0.1067	1.3628 ± 0.033	0.6145 ± 0	**0.6097 ± 0**
DBI↓	2.6657 ± 0.1529	2.5945 ± 0.1526	2.5025 ± 0.01630	2.5017 ± 0.01410	2.5017 ± 0.0200	2.9182 ± 0.3647	2.4489 ± 0	**2.442 ± 0**
LSSRI↑	−0.8626 ± 0.0911	−0.8626 ± 0.0768	−0.6551 ± 0.0027	−0.6546 ± 0.0022	−0.6586 ± 0.008	−2.4137 ± 0.079	−0.6751 ± 0	**−0.6521 ± 0**
TWI↓	1.1212 ± 0.0299	1.1213 ± 0.0254	1.0498 ± 0.0010	1.0497 ± 0.0008	1.0511 ± 0.0029	1.4638 ± 0.0095	1.057 ± 0	**1.0488 ± 0**
ACC↑	0.5182 ± 0.0962	0.5385 ± 0.0716	0.5828 ± 0.0079	0.5880 ± 0.0086	0.5729 ± 0.0206	0.4089 ± 0.0327	**0.6250 ± 0**	0.6146 ± 0
NMI↑	0.5875 ± 0.0691	0.595 ± 0.0423	0.6949 ± 0.0066	0.6968 ± 0.0042	0.6868 ± 0.0187	0.1465 ± 0.0197	**0.7386 ± 0**	0.7326 ± 0
ARI↑	0.2842 ± 0.1162	0.3082 ± 0.0899	0.3284 ± 0.0166	0.3305 ± 0.0069	0.3152 ± 0.0246	−0.0253 ± 0.0379	**0.3581 ± 0**	0.3580 ± 0

introduced. In Section 3, the proposed collaborative annealing power k-means++ clustering algorithm is delineated. In Section 4, experimental results on twelve benchmark datasets are discussed in detail. Finally, in Section 5, the conclusions are given.

2. Preliminaries

2.1. Problem formlation

Consider a dataset with n samples $\{x_i\}_{i=1}^n \in \mathbb{R}^m$ with m features to be partitioned into k clusters. k-means clustering is commonly carried out directly or indirectly by minimizing the following objective function [6]:

$$f(\Theta) = \sum_{i=1}^{n} \min_{1\le j\le k} \|x_i - \theta_j\|_2^2, \tag{1}$$

where $\theta_j \in \Re^m$ is the center (centroid) of cluster j ($j = 1, 2, \ldots, k$), $\Theta = [\theta_1, \theta_2, \ldots, \theta_k]$ is the matrix of cluster centers.

It is formulated as a minimum sum-of-squares problem in [55, 56] as follows:

$$\begin{aligned} \min_{Y,\Theta} \ & \sum_{i=1}^{n}\sum_{j=1}^{k} y_{ij}\|x_i - \theta_j\|_2^2, \\ \text{s.t.} \ & \sum_{j=1}^{k} y_{ij} = 1, i = 1, 2, \ldots, n; \\ & y_{ij} \in \{0, 1\}, \ i = 1, 2, \ldots, n; j = 1, 2, \ldots, k; \end{aligned} \tag{2}$$

where y_{ij} is the cluster indicator (i.e., $y_{ij} = 1$ implies pattern i belonging to cluster j or $y_{ij} = 0$ otherwise), for $i = 1, 2, \ldots, n; j = 1, 2, \ldots, k$. Despite its important theoretical implication, the mixed-integer optimization problem formulation above is NP-hard [57].

2.2. k-means++

k-means++ [50] is an enhanced clustering algorithm by generating initial centers randomly in favor of high dissimilarity between selected initial centers. Starting with an arbitrary sample as the first center, a subsequent center is selected from the samples with a probability proportional to the distance between a sample and its nearest existing center:

$$P(x) = \frac{d(x)^2}{\sum_{x\in\mathcal{X}} d(x)^2}, \tag{3}$$

where x is a sample, $\mathcal{X}$ is a set of samples other than selected centers, $d(x)$ is the distance between x and its nearest center.

Table 5
The mean values and standard deviations of internal and external cluster validity indices resulting from CAPKM++ and seven baselines on ORL10P and WarpPIE10P.

ORL10P	KM	KM++	PKM	PKM++	EWPKM	SC	HC	CAPKM++
WGSS↓	2.7083 ± 0.1705	2.625 ± 0.1315	2.4151 ± 0	2.4130 ± 0.0061	2.4185 ± 0.0129	5.7656 ± 0.1181	2.4089 ± 0	**2.3809 ± 0**
MRI↓	0.6786 ± 0.0275	0.6682 ± 0.0169	0.6489 ± 0	0.6466 ± 0.0055	0.6493 ± 0.0018	1.0030 ± 0.0225	0.6493 ± 0	**0.635 ± 0**
GPI↓	0.0182 ± 0.0087	0.0154 ± 0.0048	0.0122 ± 0	0.0117 ± 0.0012	0.0123 ± 0.0002	0.2362 ± 0.019	0.0115 ± 0	**0.0089 ± 0**
BHGI↑	0.8353 ± 0.069	0.8536 ± 0.0444	0.8601 ± 0	0.8655 ± 0.0135	0.8595 ± 0.0019	−0.0097 ± 0.0692	0.8863 ± 0	**0.9058 ± 0**
CI↓	0.103 ± 0.0373	0.0927 ± 0.0246	0.0888 ± 0	0.0859 ± 0.0073	0.0893 ± 0.0018	0.5074 ± 0.0363	0.0711 ± 0	**0.0599 ± 0**
TI↑	0.3929 ± 0.0477	0.3931 ± 0.0351	0.3596 ± 0	0.3616 ± 0.005	0.3596 ± 0.0008	−0.0065 ± 0.0478	**0.3983 ± 0**	0.3931 ± 0
DGI↑	1.1115 ± 0.1433	1.1548 ± 0.1582	1.1878 ± 0	1.1972 ± 0.0308	1.1937 ± 0.0239	0.5744 ± 0.0931	1.3129 ± 0	**1.3233 ± 0**
RLI↑	0.2326 ± 0.0059	0.2343 ± 0.0054	0.2395 ± 0	0.2395 ± 0.0002	0.2393 ± 0.0004	0.0956 ± 0.0057	0.2394 ± 0	**0.2400 ± 0**
CHI↑	1.2933 ± 0.1212	1.3502 ± 0.1114	1.51 ± 0	1.5122 ± 0.0063	1.5066 ± 0.0131	0.1003 ± 0.0132	1.5165 ± 0	**1.5460 ± 0**
RTI↓	1.5366 ± 0.4225	1.3206 ± 0.298	1.0508 ± 0	1.0531 ± 0.0147	1.0585 ± 0.0324	14.3179 ± 9.8311	0.8902 ± 0	**0.8424 ± 0**
WGI↑	0.3230 ± 0.0271	0.3386 ± 0.0254	0.39 ± 0	0.387 ± 0.0074	0.3891 ± 0.0034	0.0532 ± 0.0128	0.3914 ± 0	**0.3944 ± 0**
DI↑	0.4177 ± 0.0627	0.4336 ± 0.0582	0.4534 ± 0	0.4554 ± 0.0091	0.4557 ± 0.0091	0.2199 ± 0.026	0.4941 ± 0	**0.4941 ± 0**
BHI↑	0.0249 ± 0.0015	0.0238 ± 0.0014	0.0227 ± 0	0.0229 ± 0.0004	0.0227 ± 0.0001	**0.0273 ± 0.0069**	0.0219 ± 0	0.0228 ± 0
PBMI↑	0.0038 ± 0.0004	0.0038 ± 0.0004	0.0042 ± 0	0.0042 ± 0	**0.0042 ± 0**	0.002 ± 0.0004	0.0037 ± 0	0.0039 ± 0
XBI↓	1.1521 ± 0.2689	1.0856 ± 0.3352	0.8864 ± 0	0.8786 ± 0.0316	0.8797 ± 0.0308	4.8812 ± 1.0998	0.7443 ± 0	**0.7356 ± 0**
DBI↓	1.7443 ± 0.1419	1.64 ± 0.1006	1.4956 ± 0	1.5109 ± 0.0382	1.4995 ± 0.0128	3.2496 ± 0.7087	**1.4603 ± 0**	1.4798 ± 0
LSSRI↑	0.2143 ± 0.1138	0.2713 ± 0.0882	0.4121 ± 0	0.4136 ± 0.0042	0.4098 ± 0.0088	−2.3071 ± 0.1274	0.4164 ± 0	**0.4357 ± 0**
TWI↓	0.2708 ± 0.017	0.2623 ± 0.0132	0.2415 ± 0	0.2413 ± 0.0006	0.2418 ± 0.0013	0.551 ± 0.0066	0.2409 ± 0	**0.2381 ± 0**
ACC↑	0.6450 ± 0.0871	0.6980 ± 0.0766	0.8100 ± 0	**0.8105 ± 0.0076**	0.809 ± 0.0165	0.202 ± 0.0235	0.7700 ± 0	0.7900 ± 0
NMI↑	0.7722 ± 0.0505	0.8073 ± 0.0413	0.8584 ± 0	0.8587 ± 0.0085	0.8604 ± 0.0102	0.1898 ± 0.0252	0.8678 ± 0	**0.8801 ± 0**
ARI↑	0.5369 ± 0.0985	0.5983 ± 0.084	0.7161 ± 0	0.7172 ± 0.0144	0.7191 ± 0.0174	−0.0002 ± 0.0075	0.684 ± 0	**0.7273 ± 0**
WarpPIE10P	**KM**	**KM++**	**PKM**	**PKM++**	**EWPKM**	**SC**	**HC**	**CAPKM++**
WGSS↓	3.6075 ± 0.0994	3.6368 ± 0.1163	3.4688 ± 0.0093	3.4648 ± 0	3.465 ± 0	8.4389 ± 0.2204	3.6281 ± 0	**3.4481 ± 0**
MRI↓	0.5535 ± 0.0137	0.5536 ± 0.0112	0.5517 ± 0.0022	0.5525 ± 0	0.5529 ± 0	1.0093 ± 0.0073	0.5548 ± 0	**0.5368 ± 0**
GPI↓	0.0164 ± 0.0031	0.0167 ± 0.0029	0.0151 ± 0.0004	0.0153 ± 0	0.0154 ± 0	0.2318 ± 0.0090	0.0168 ± 0	**0.0131 ± 0**
BHGI↑	0.8388 ± 0.0268	0.8438 ± 0.0191	0.8337 ± 0.0034	0.8324 ± 0	0.8312 ± 0	−0.0044 ± 0.0150	0.826 ± 0	**0.8594 ± 0**
CI↓	0.088 ± 0.0124	0.0844 ± 0.0083	0.0937 ± 0.0016	0.0943 ± 0	0.0946 ± 0	0.4982 ± 0.0117	0.0925 ± 0	**0.0788 ± 0**
TI↑	0.3776 ± 0.0165	**0.3898 ± 0.0200**	0.3555 ± 0.0008	0.3554 ± 0	0.3549 ± 0	−0.0031 ± 0.0104	0.363 ± 0	0.3712 ± 0
DGI↑	0.4701 ± 0.0932	0.4836 ± 0.0817	0.4254 ± 0.0337	0.4401 ± 0	0.4401 ± 0	**0.6526 ± 0.0124**	0.6178 ± 0	0.6128 ± 0
RLI↑	0.2551 ± 0.0020	0.2550 ± 0.0018	0.2575 ± 0.0003	**0.2576 ± 0**	0.2575 ± 0	0.1484 ± 0.0062	0.2505 ± 0	0.2575 ± 0
CHI↑	2.0493 ± 0.0805	2.0255 ± 0.0942	2.1691 ± 0.0084	2.1727 ± 0	2.1725 ± 0	0.3034 ± 0.0322	2.0299 ± 0	**2.1881 ± 0**
RTI↓	1.1029 ± 0.1524	1.0443 ± 0.173	0.999 ± 0.0501	0.9746 ± 0	0.9747 ± 0	3.1805 ± 0.0831	**0.9082 ± 0**	0.9166 ± 0
WGI↑	0.303 ± 0.0210	0.308 ± 0.0131	0.3029 ± 0.0016	0.3037 ± 0	0.3038 ± 0	0.1619 ± 0.0164	0.3058 ± 0	**0.3298 ± 0**
DI↑	0.1647 ± 0.0374	0.1606 ± 0.0284	0.1594 ± 0.0131	0.165 ± 0	0.165 ± 0	0.1653 ± 0	0.2062 ± 0	**0.2357 ± 0**
BHI↑	0.0171 ± 0.0006	0.0172 ± 0.0006	0.0164 ± 0.0003	0.0163 ± 0	0.0163 ± 0	0.0120 ± 0.0006	**0.0174 ± 0**	0.0170 ± 0
PBMI↑	**0.0065 ± 0.0006**	0.0064 ± 0.0007	0.0064 ± 0.0001	0.0064 ± 0	0.0063 ± 0	0.0031 ± 0.0001	0.0063 ± 0	0.0063 ± 0
XBI↓	6.3067 ± 2.9265	5.9085 ± 2.2502	6.5469 ± 1.4846	5.9511 ± 0	5.9515 ± 0	3.3962 ± 0.0887	3.0665 ± 0	**2.9986 ± 0**
DBI↓	1.6448 ± 0.0754	1.6389 ± 0.0794	1.5938 ± 0.0347	1.5788 ± 0	1.5765 ± 0	1.6771 ± 0.067	**1.5338 ± 0**	1.6021 ± 0
LSSRI↑	0.7167 ± 0.0406	0.7048 ± 0.0475	0.7743 ± 0.0039	0.7760 ± 0	0.7759 ± 0	−1.1990 ± 0.1217	0.708 ± 0	**0.7830 ± 0**
TWI↓	0.3608 ± 0.0099	0.3637 ± 0.0116	0.3469 ± 0.0009	0.3465 ± 0	0.3465 ± 0	0.8439 ± 0.0220	0.3628 ± 0	**0.3448 ± 0**
ACC↑	0.2795 ± 0.0167	**0.289 ± 0.0237**	0.2719 ± 0.0034	0.2714 ± 0	0.2714 ± 0	0.2879 ± 0.0110	0.2857 ± 0	0.2810 ± 0
NMI↑	0.2957 ± 0.0309	0.2978 ± 0.0419	0.3048 ± 0.0065	0.3061 ± 0	0.3141 ± 0	**0.3837 ± 0.0099**	0.3614 ± 0	0.3140 ± 0
ARI↑	0.0869 ± 0.0209	0.0856 ± 0.0229	0.0922 ± 0.0058	0.0918 ± 0	0.0959 ± 0	0.0903 ± 0.0177	**0.1117 ± 0**	0.0940 ± 0

2.3. Power k-means clustering

In [48], the power k-means algorithm is proposed to improve k-means algorithms by minimizing an annealed power function and demonstrated to perform better than Lloyd's algorithm [6] and k-harmonic means [7]. The power mean is defined as:

$$M_s(y) = M_s(y_1, y_2, \ldots, y_k) = \left(\frac{1}{k}\sum_{i=1}^{k} y_i^s\right)^{\frac{1}{s}}, \tag{4}$$

where s is the exponent.

Let $f_s(\Theta) := \sum_{i=1}^{n} M_s(\|x_i - \theta_1\|_2^2, \ldots, \|x_i - \theta_k\|_2^2)$. As $\lim_{s\to-\infty} M_s(y) = \min_{1\le j\le k}\{y_1, \ldots, y_k\}$ [58],

$$\lim_{s\to-\infty}\sum_{i=1}^{n} M_s(\|x_i-\theta_1\|^2, \ldots, \|x_i-\theta_k\|^2) = \sum_{i=1}^{n}\min_{1\le j\le k}\|x_i-\theta_j\|_2^2;$$

i.e., $\lim_{s\to-\infty} f_s(\Theta) = f(\Theta)$.

If $s < 1$, $M_s(y)$ is concave [48]; i.e., at any anchor point z,

$$M_s(y) \le M_s(z) + \sum_{j=1}^{k}\frac{\partial}{\partial z_j}M_s(z)(y_j - z_j). \tag{5}$$

The partial derivative of $M_s(z)$ with respect to z_j is derived as follows:

$$\frac{\partial}{\partial z_j}M_s(z) = \left(\frac{1}{k}\sum_{i=1}^{k} z_i^s\right)^{\frac{1}{s}-1}\frac{1}{k}z_j^{s-1}. \tag{6}$$

By substituting $y_{ij} = \|x_i - \theta_j(t+1)\|_2^2$, $z_{ij} = \|x_i - \theta_j(t)\|_2^2$, and the partial derivative in (6) into inequality (5) and summing it from $i = 1, \ldots, n$, we have the following inequality:

$$f_s(\Theta(t+1)) \le \overbrace{f_s(\Theta(t)) - \sum_{i=1}^{n}\sum_{j=1}^{k} w_{ij}(t)\left\|x_i - \theta_j(t)\right\|_2^2}^{\text{Independent of } \theta_j(t+1)(j=1,2,\ldots,k)} + \sum_{i=1}^{n}\sum_{j=1}^{k} w_{ij}(t)\|x_i - \theta_j(t+1)\|_2^2, \tag{7}$$

where t is the iteration index and $w_{ij}(t) \in [0, 1]$ is the weight for sample i to be assigned to cluster j at iteration t ($i = 1, \ldots, n$; $j = 1, \ldots, k$) as defined below:

$$w_{ij}(t) = \frac{\frac{1}{k}\|x_i - \theta_j(t)\|_2^{2(s-1)}}{\left(\frac{1}{k}\sum_{l=1}^{k}\|x_i - \theta_l(t)\|_2^{2s}\right)^{1-\frac{1}{s}}}. \tag{8}$$

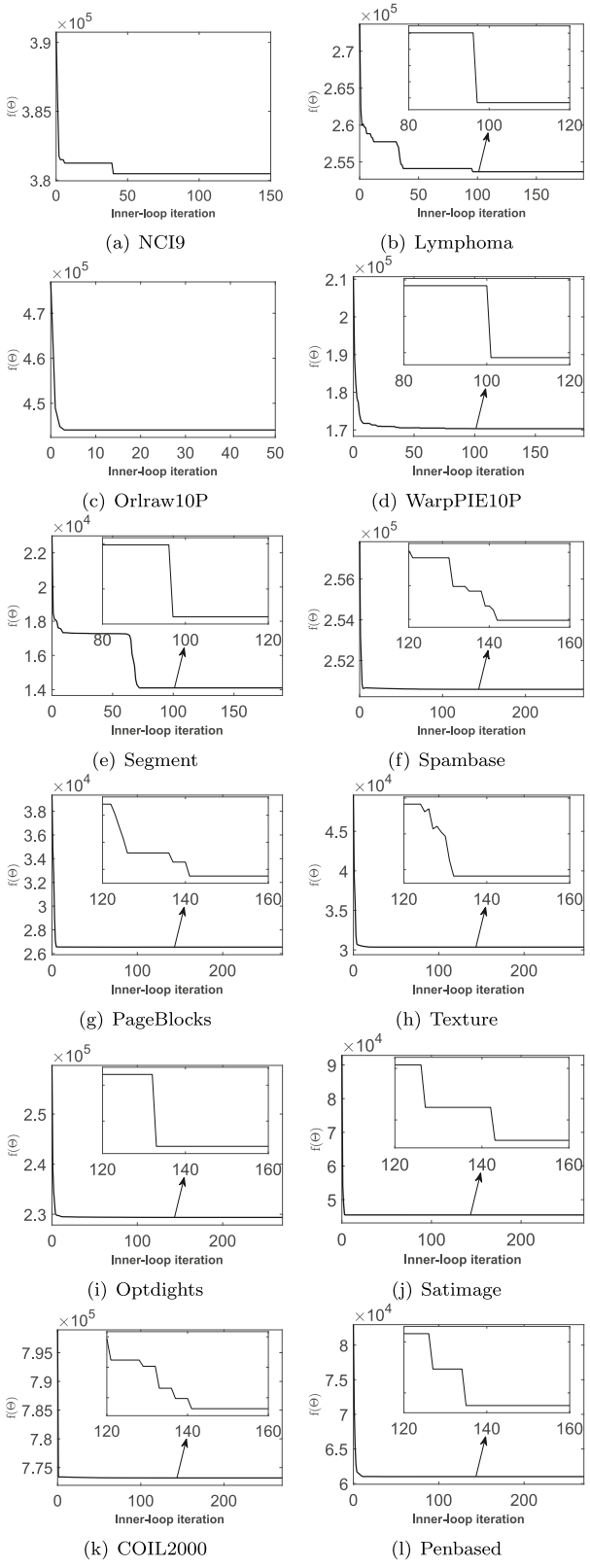

Fig. 3. Snapshots of the objective function values in (1) value in the inner loop (Steps 4–8) of CAPKM++ on the twelve datasets.

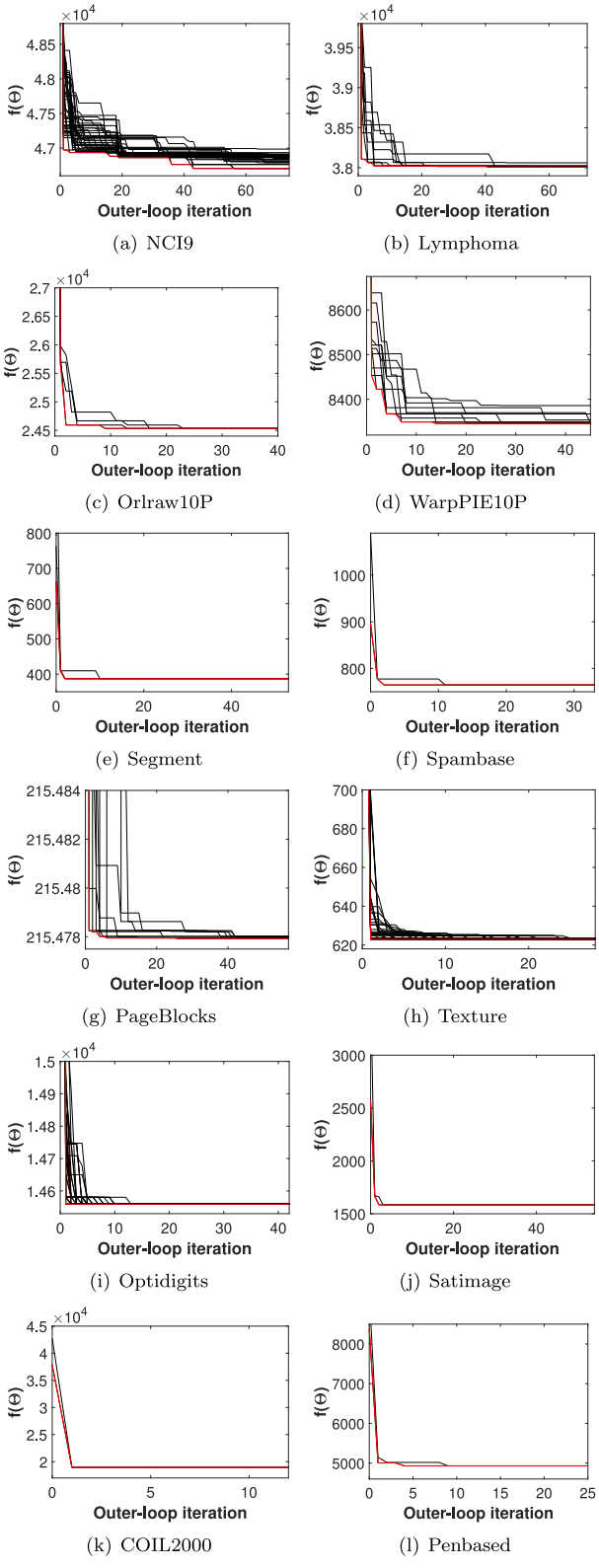

Fig. 4. The convergent behavior of the CAPKM++ algorithm on the twelve datasets.

A majorization function is defined by the last term on the right-hand side of (7). By zeroing the partial derivative of the majorization function with respect to $\theta_j(t+1)$ for $j = 1, \ldots, k$ to minimize $f_s(\Theta)$, the cluster center updating rule is derived as

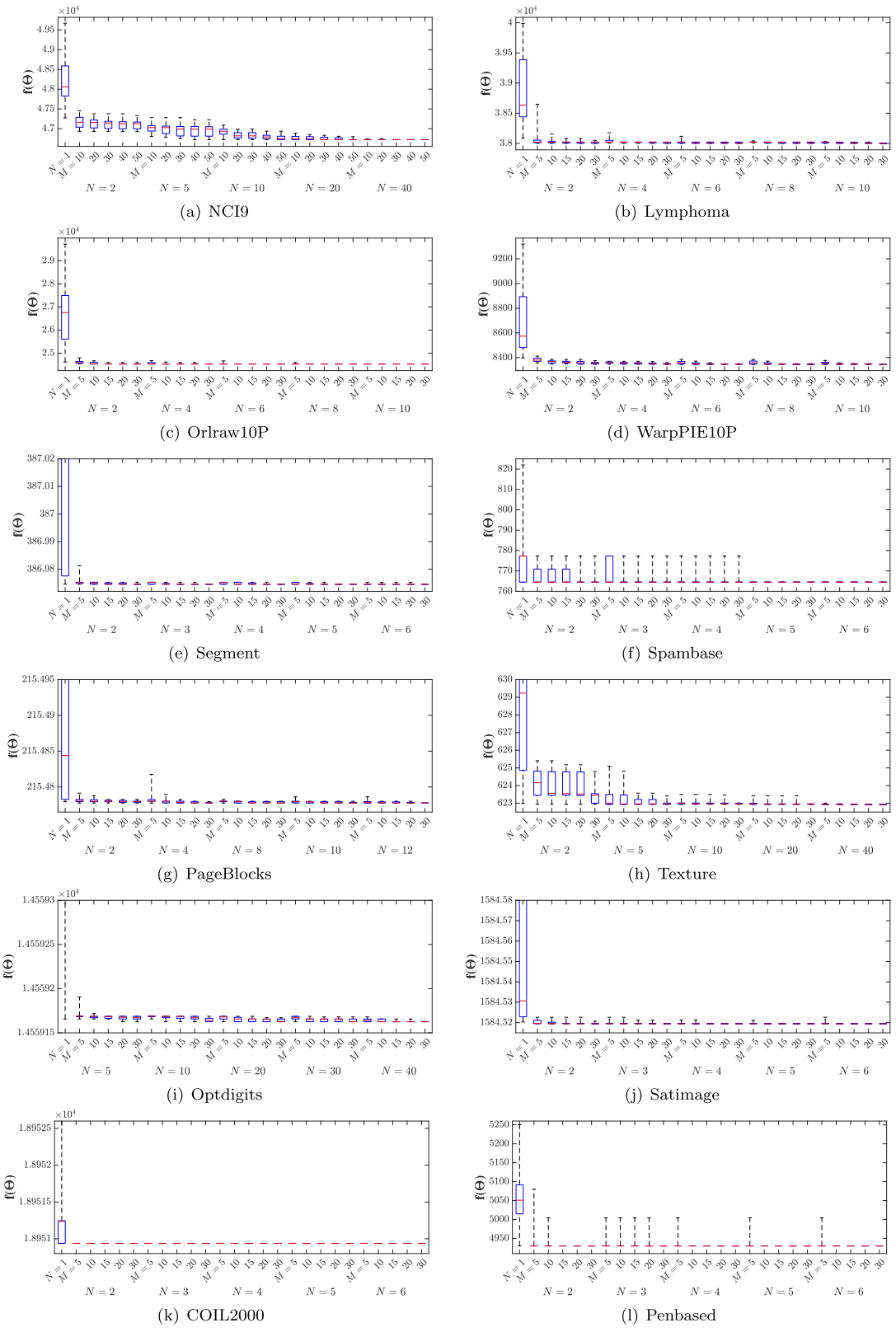

Fig. 5. Monte Carlo test results of the CAPKM++ algorithm with several values of M and N on the twelve datasets.

follows:

$$\theta_j(t+1) = \frac{1}{\sum_{i=1}^{n} w_{ij}(t)} \sum_{i=1}^{n} w_{ij}(t) x_i. \tag{9}$$

The exponent s is reduced iteratively toward $-\infty$, according to a cooling schedule: From $s(0) < 0$, for $t = 0, 1, \ldots$;

$$s(t+1) = \eta s(t), \tag{10}$$

where $\eta > 1$.

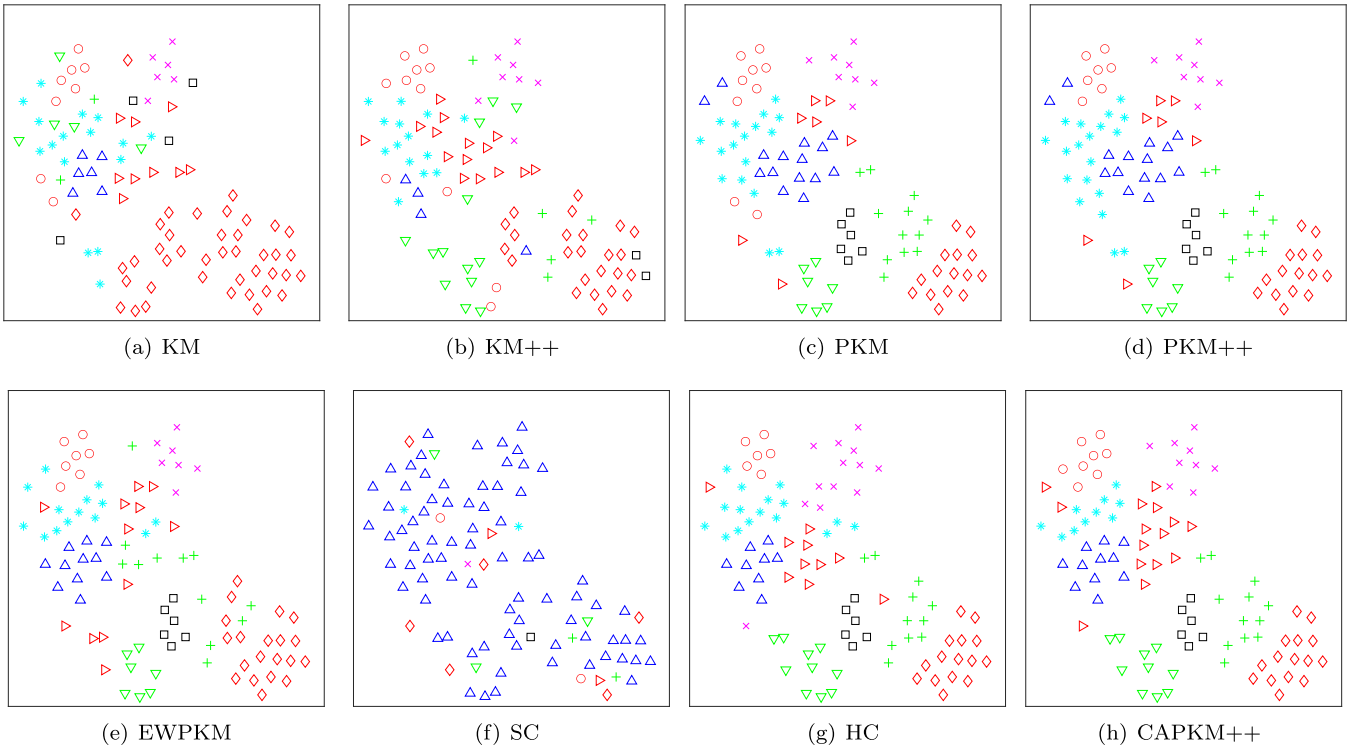

Fig. 6. The clustering results of different clustering algorithms on Lymphoma mapped to planes using the t-SNE algorithm.

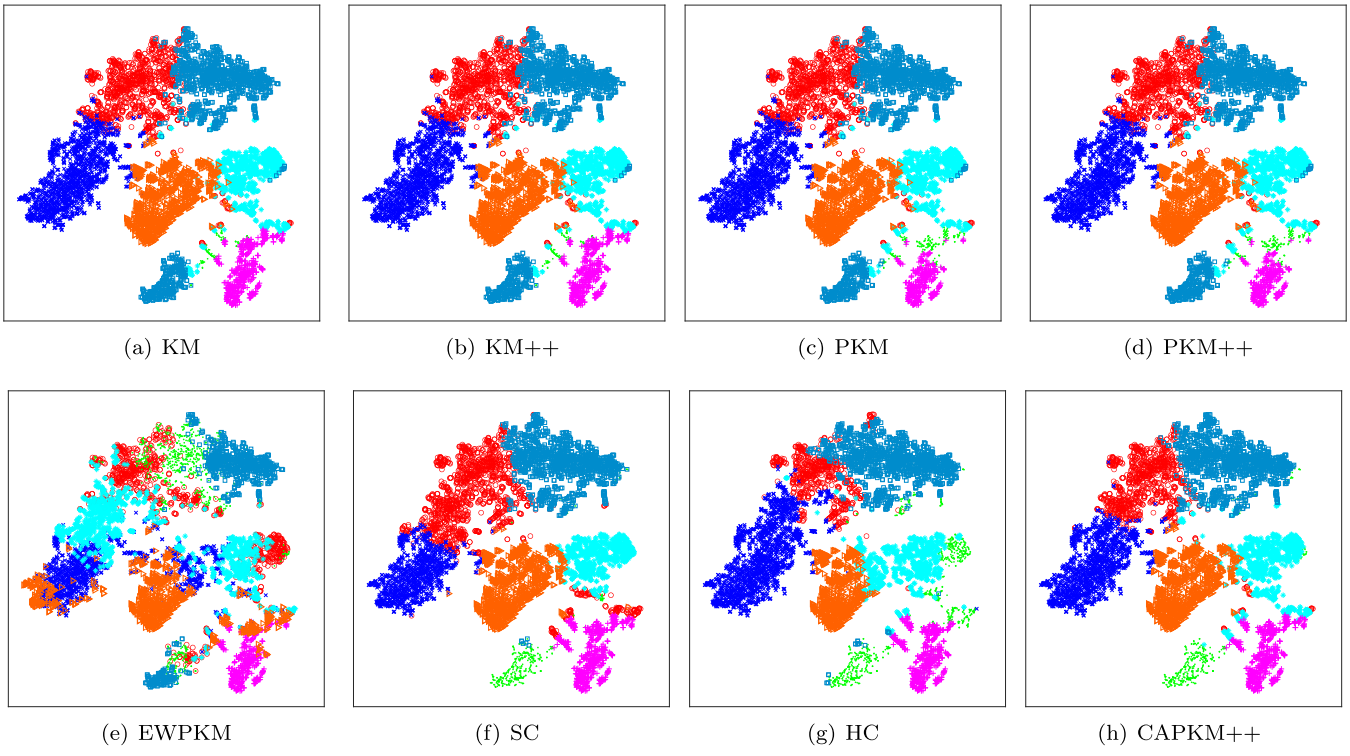

Fig. 7. The clustering results of different clustering algorithms on Satimage mapped to planes using the t-SNE algorithm.

It is worth mentioning that the clustering results by Power k-means are not globally optimal since the resulting clusters depend heavily on the anchor points where the majorization functions locate.

2.4. Collaborative neurodynamic optimization

In global optimization, initial solutions are crucially important. A hybrid intelligence framework, called collaborative

Table 6
The mean values and standard deviations of internal and external cluster validity indices resulting from CAPKM++ and seven baselines on Segment and SpamBase.

Segment	KM	KM++	PKM	PKM++	EWPKM	SC	HC	CAPKM++
WGSS↓	23.1881 ± 1.2563	22.9184 ± 0.9713	22.0185 ± 0.5957	22.2194 ± 0.6081	25.1188 ± 0.596	30.0177 ± 0	23.3898 ± 0	**21.4986 ± 0**
MRI↓	0.4459 ± 0.0151	0.4430 ± 0.0129	0.4334 ± 0.0109	0.4383 ± 0.0124	0.4628 ± 0.008	0.5196 ± 0	0.4548 ± 0	**0.4243 ± 0**
GPI↓	0.0215 ± 0.0038	0.0208 ± 0.0031	0.0188 ± 0.0025	0.0200 ± 0.003	0.0258 ± 0.0023	0.0469 ± 0	0.0238 ± 0	**0.0168 ± 0**
BHGI↑	0.8516 ± 0.0156	0.8540 ± 0.0132	0.8558 ± 0.012	0.8497 ± 0.0149	0.8106 ± 0.0078	0.7666 ± 0	0.8152 ± 0	**0.8634 ± 0**
CI↓	0.0613 ± 0.0057	0.0605 ± 0.0048	0.0596 ± 0.0045	0.0617 ± 0.0054	0.0775 ± 0.0028	0.0957 ± 0	0.0751 ± 0	**0.0565 ± 0**
TI↑	0.4572 ± 0.0272	0.4544 ± 0.0220	0.4357 ± 0.0093	0.4369 ± 0.0085	0.4225 ± 0.0072	**0.4858 ± 0**	0.4136 ± 0	0.4289 ± 0
DGI↑	0.0875 ± 0.0326	0.0864 ± 0.0271	0.0991 ± 0.0209	0.096 ± 0.0176	0.1012 ± 0.0121	**0.2905 ± 0**	0.2132 ± 0	0.108 ± 0
RLI↑	0.286 ± 0.0073	**0.2889 ± 0.0088**	0.2841 ± 0.0015	0.2838 ± 0.0014	0.2821 ± 0.0012	0.2718 ± 0	0.283 ± 0	0.2853 ± 0
CHI↑	3.3835 ± 0.2276	3.4308 ± 0.1877	3.6073 ± 0.1238	3.5657 ± 0.1262	3.0380 ± 0.0949	2.3772 ± 0	3.3341 ± 0	**3.7154 ± 0**
RTI↓	0.6924 ± 0.0642	0.6967 ± 0.0992	0.6302 ± 0.0505	0.6528 ± 0.0555	2.1570 ± 1.5130	1.8239 ± 0	0.7898 ± 0	**0.5894 ± 0**
WGI↑	0.4339 ± 0.0190	0.4374 ± 0.021	0.4432 ± 0.0180	0.4394 ± 0.0166	0.3457 ± 0.0212	0.306 ± 0	0.4123 ± 0	**0.4582 ± 0**
DI↑	0.0152 ± 0.0044	0.0159 ± 0.0044	0.0151 ± 0.0033	0.0149 ± 0.0027	0.0161 ± 0.0017	**0.0453 ± 0**	0.0340 ± 0	0.0159 ± 0
BHI↑	0.0101 ± 0.0016	0.0102 ± 0.0015	0.0093 ± 0.0002	0.0092 ± 0.0002	**0.0105 ± 0.0003**	0.0087 ± 0	0.0093 ± 0	0.0094 ± 0
PBMI↑	0.0226 ± 0.0011	0.0226 ± 0.0012	0.0238 ± 0.0006	0.0237 ± 0.0005	0.0210 ± 0.0006	0.0193 ± 0	0.0226 ± 0	**0.0241 ± 0**
XBI↓	243.722 ± 124.8857	237.7285 ± 156.1841	245.4361 ± 198.2515	243.7731 ± 196.4317	195.8641 ± 73.9913	**20.4288 ± 0**	35.7993 ± 0	172.0822 ± 0
DBI↓	1.2732 ± 0.0777	1.2497 ± 0.0890	1.2417 ± 0.0686	1.2466 ± 0.0613	1.7887 ± 0.2632	1.5252 ± 0	1.2809 ± 0	**1.1902 ± 0**
LSSRI↑	1.2167 ± 0.0692	1.2314 ± 0.0548	1.2824 ± 0.0345	1.2708 ± 0.0352	1.1107 ± 0.0315	0.8659 ± 0	1.2042 ± 0	**1.3125 ± 0**
TWI↓	3.3126 ± 0.1795	3.2741 ± 0.1388	3.1455 ± 0.0851	3.1742 ± 0.0869	3.5884 ± 0.0851	4.2882 ± 0	3.3414 ± 0	**3.0712 ± 0**
ACC↑	0.5793 ± 0.0656	0.5931 ± 0.051	0.6246 ± 0.0523	0.6214 ± 0.0449	0.5848 ± 0.0378	0.4848 ± 0	**0.7078 ± 0**	0.6658 ± 0
NMI↑	0.6072 ± 0.0138	0.6093 ± 0.0152	0.6066 ± 0.0132	0.6115 ± 0.0128	0.6251 ± 0.0172	**0.6626 ± 0**	0.6377 ± 0	0.6124 ± 0
ARI↑	0.4594 ± 0.0425	0.4677 ± 0.0345	0.4795 ± 0.0344	0.4803 ± 0.0293	0.473 ± 0.0431	0.404 ± 0	**0.5382 ± 0**	0.5045 ± 0
SpamBase	KM	KM++	PKM	PKM++	EWPKM	SC	HC	CAPKM++
WGSS↓	13.6707 ± 0.2724	13.6938 ± 0.3615	13.6374 ± 0	13.6374 ± 0	14.0353 ± 0	14.1235 ± 0	13.4150 ± 0	**13.4122 ± 0**
MRI↓	0.7574 ± 0.2370	0.6672 ± 0.2597	0.8998 ± 0	0.8998 ± 0	0.8797 ± 0	0.4893 ± 0	0.3282 ± 0	**0.3236 ± 0**
GPI↓	0.1439 ± 0.0949	0.1051 ± 0.1022	0.2047 ± 0	0.2047 ± 0	0.1822 ± 0	0.0015 ± 0	0.0006 ± 0	**0.0005 ± 0**
BHGI↑	0.3853 ± 0.3314	0.5208 ± 0.3631	0.1793 ± 0	0.1793 ± 0	0.2167 ± 0	0.9027 ± 0	0.9584 ± 0	**0.9646 ± 0**
CI↓	0.3747 ± 0.1473	0.3439 ± 0.1796	0.436 ± 0	0.4360 ± 0	0.456 ± 0	0.4021 ± 0	0.0981 ± 0	**0.0913 ± 0**
TI↑	0.1351 ± 0.0140	0.1426 ± 0.0476	0.1266 ± 0	0.1266 ± 0	0.1478 ± 0	0.1579 ± 0	**0.1630 ± 0**	0.1617 ± 0
DGI↑	0.2464 ± 0.2864	0.3215 ± 0.3411	0.0450 ± 0	0.0450 ± 0	0.1827 ± 0	**1.9618 ± 0**	0.6690 ± 0	0.6529 ± 0
RLI↑	0.1468 ± 0.0267	0.1477 ± 0.0371	0.1398 ± 0	0.1398 ± 0	0.1435 ± 0	0.0918 ± 0	0.1936 ± 0	**0.1938 ± 0**
CHI↑	0.0792 ± 0.0207	0.0777 ± 0.0271	0.0814 ± 0	0.0814 ± 0	0.0508 ± 0	0.0442 ± 0	0.0994 ± 0	**0.0996 ± 0**
RTI↓	2.1021 ± 1.2761	1.5711 ± 1.3873	2.9158 ± 0	2.9158 ± 0	3.6234 ± 0	0.1758 ± 0	0.0739 ± 0	**0.0716 ± 0**
WGI↑	0.3699 ± 0.2297	0.4591 ± 0.2512	0.229 ± 0	0.229 ± 0	0.1957 ± 0	0.6869 ± 0	0.7756 ± 0	**0.7788 ± 0**
DI↑	0.0315 ± 0.0404	0.0426 ± 0.0454	0.0042 ± 0	0.0042 ± 0	0.0194 ± 0	**0.1851 ± 0**	0.1032 ± 0	0.0989 ± 0
BHI↑	0.0033 ± 0.0006	0.0035 ± 0.0009	0.0029 ± 0	0.0029 ± 0	0.0031 ± 0	0.0016 ± 0	**0.0045 ± 0**	0.0044 ± 0
PBMI↑	0.0025 ± 0.0042	0.0038 ± 0.0048	0.0003 ± 0	0.0003 ± 0	0.0002 ± 0	0.0046 ± 0	0.0103 ± 0	**0.0106 ± 0**
XBI↓	494.4740 ± 420.8625	402.0681 ± 452.6952	1288.1196 ± 0	1288.1196 ± 0	66.0743 ± 0	**0.6541 ± 0**	2.8644 ± 0	3.1153 ± 0
DBI↓	2.2358 ± 0.9665	1.8316 ± 1.0813	2.8405 ± 0	2.8405 ± 0	3.2839 ± 0	**0.3808 ± 0**	0.5975 ± 0	0.5818 ± 0
LSSRI↑	−2.5957 ± 0.4157	−2.7453 ± 0.8702	−2.508 ± 0	−2.508 ± 0	−2.9805 ± 0	−3.1189 ± 0	−2.309 ± 0	**−2.3067 ± 0**
TWI↓	6.8354 ± 0.1362	6.8469 ± 0.1808	6.8187 ± 0	6.8187 ± 0	7.0177 ± 0	7.0618 ± 0	6.7075 ± 0	**6.7061 ± 0**
ACC↑	0.7403 ± 0.0927	0.6963 ± 0.1066	0.8001 ± 0	0.8001 ± 0	**0.8162 ± 0**	0.598 ± 0	0.5984 ± 0	0.5987 ± 0
NMI↑	0.1874 ± 0.1142	0.1446 ± 0.1195	0.2617 ± 0	0.2617 ± 0	**0.3455 ± 0**	0.0225 ± 0	0.0218 ± 0	0.0214 ± 0
ARI↑	0.2495 ± 0.1685	0.1751 ± 0.1868	0.3586 ± 0	0.3586 ± 0	**0.3926 ± 0**	−0.0052 ± 0	−0.0049 ± 0	−0.0048 ± 0

Table 7
The mean values and standard deviations of internal and external cluster validity indices resulting from CAPKM++ and seven baselines on PageBlocks and Texture.

PageBlocks	KM	KM++	PKM	PKM++	EWPKM	SC	HC	CAPKM++
WGSS↓	21.8359 ± 0.5068	22.1510 ± 0.6750	21.6410 ± 0.0478	21.6002 ± 0.0594	42.8807 ± 0	27.9910 ± 0	23.7374 ± 0	**21.5478 ± 0**
MRI↓	0.4071 ± 0.006	0.4082 ± 0.0069	0.4069 ± 0.0049	0.4027 ± 0.0061	0.5267 ± 0	0.4738 ± 0	0.4167 ± 0	**0.3973 ± 0**
GPI↓	0.0505 ± 0.0022	0.0516 ± 0.003	0.0501 ± 0.0013	0.049 ± 0.0016	0.0896 ± 0	0.0797 ± 0	0.0547 ± 0	**0.0475 ± 0**
BHGI↑	0.7601 ± 0.0170	0.7638 ± 0.0156	0.7566 ± 0.0155	0.7698 ± 0.0193	0.6192 ± 0	0.6547 ± 0	0.7583 ± 0	**0.7871 ± 0**
CI↓	0.0836 ± 0.0055	0.0833 ± 0.0055	0.0842 ± 0.0050	0.0799 ± 0.0062	0.2049 ± 0	0.1327 ± 0	0.0889 ± 0	**0.0744 ± 0**
TI↑	0.4941 ± 0.0237	0.5053 ± 0.0224	0.4859 ± 0.0202	0.5032 ± 0.0251	0.4248 ± 0	0.4450 ± 0	0.5102 ± 0	**0.5257 ± 0**
DGI↑	0.0265 ± 0.0069	0.0267 ± 0.0067	0.0324 ± 0.0002	0.0325 ± 0.0002	0.0098 ± 0	0.0096 ± 0	**0.061 ± 0**	0.0327 ± 0
RLI↑	0.2584 ± 0.0040	0.2654 ± 0.0142	0.2568 ± 0.0027	0.2591 ± 0.0034	**0.3118 ± 0**	0.2491 ± 0	0.2541 ± 0	0.2622 ± 0
CHI↑	2.0719 ± 0.0682	2.0294 ± 0.0912	2.0981 ± 0.0069	2.1039 ± 0.0085	0.5635 ± 0	1.3952 ± 0	1.8244 ± 0	**2.1114 ± 0**
RTI↓	0.6501 ± 0.184	0.5632 ± 0.1742	0.7193 ± 0.1552	0.5866 ± 0.1927	0.7699 ± 0	1.4506 ± 0	0.5442 ± 0	**0.4160 ± 0**
WGI↑	0.4656 ± 0.0261	0.4806 ± 0.0266	0.4558 ± 0.0214	0.4741 ± 0.0266	0.429 ± 0	0.425 ± 0	0.4269 ± 0	**0.4978 ± 0**
DI↑	0.0027 ± 0.0008	0.0033 ± 0.0011	0.0032 ± 0.0001	0.0033 ± 0.0001	0.0025 ± 0	0.0010 ± 0	**0.0067 ± 0**	0.0034 ± 0
BHI↑	0.0062 ± 0.0008	0.0074 ± 0.002	0.0059 ± 0.0005	0.0063 ± 0.0006	**0.0217 ± 0**	0.0055 ± 0	0.0059 ± 0	0.0068 ± 0
PBMI↑	0.0110 ± 0.0014	0.0131 ± 0.0062	0.0110 ± 0.0011	0.0120 ± 0.0014	0.0093 ± 0	0.0111 ± 0	0.0106 ± 0	**0.0132 ± 0**
XBI↓	1582.8607 ± 1012.0129	1219.3413 ± 902.5958	848.1899 ± 101.0761	761.9876 ± 125.7117	3453.5833 ± 0	9716.1102 ± 0	**261.8155 ± 0**	651.1548 ± 0
DBI↓	1.0210 ± 0.0261	**1.0073 ± 0.0669**	1.0240 ± 0.0015	1.0226 ± 0.0017	1.5355 ± 0	1.0918 ± 0	1.0082 ± 0	1.0206 ± 0
LSSRI↑	0.7279 ± 0.0340	0.7067 ± 0.0454	0.741 ± 0.0033	0.7438 ± 0.0041	−0.5736 ± 0	0.3331 ± 0	0.6013 ± 0	**0.7474 ± 0**
TWI↓	4.3672 ± 0.1014	4.4302 ± 0.1350	4.3282 ± 0.0096	4.32 ± 0.0119	8.5761 ± 0	5.5982 ± 0	4.7475 ± 0	**4.3096 ± 0**
ACC↑	0.4549 ± 0.0622	0.4882 ± 0.0619	0.4325 ± 0.0496	0.4748 ± 0.0618	**0.7178 ± 0**	0.4793 ± 0	0.5387 ± 0	0.5302 ± 0
NMI↑	0.1536 ± 0.0075	0.158 ± 0.0089	0.1522 ± 0.0010	0.1514 ± 0.0012	0.0972 ± 0	**0.1737 ± 0**	0.1578 ± 0	0.1506 ± 0
ARI↑	0.0978 ± 0.0104	0.1046 ± 0.0114	0.0933 ± 0.0074	0.0996 ± 0.0091	0.0570 ± 0	0.0866 ± 0	**0.1226 ± 0**	0.1081 ± 0
Texture	KM	KM++	PKM	PKM++	EWPKM	SC	HC	CAPKM++
WGSS↓	16.2806 ± 0.8137	15.9375 ± 0.5631	15.6271 ± 0.0031	15.6107 ± 0.02	15.7883 ± 0.0003	21.6214 ± 0.822	17.3806 ± 0	**15.5733 ± 0**
MRI↓	0.3414 ± 0.0114	0.3378 ± 0.0081	0.3339 ± 0.0001	0.3334 ± 0.0005	0.3343 ± 0	0.4109 ± 0.0082	0.3535 ± 0	**0.3329 ± 0**
GPI↓	0.0094 ± 0.0018	0.0088 ± 0.0013	0.0082 ± 0	0.0081 ± 0.0001	0.0082 ± 0	0.024 ± 0.0019	0.0115 ± 0	**0.0079 ± 0**
BHGI↑	0.9023 ± 0.0134	0.9099 ± 0.0116	0.9138 ± 0.0001	0.9145 ± 0.0009	0.9134 ± 0	0.7988 ± 0.0122	0.8921 ± 0	**0.9202 ± 0**
CI↓	0.0364 ± 0.0047	0.0334 ± 0.0040	0.0323 ± 0.0001	0.0321 ± 0.0003	0.0327 ± 0	0.0699 ± 0.0047	0.0389 ± 0	**0.0299 ± 0**
TI↑	0.3944 ± 0.0098	0.4015 ± 0.0150	0.3974 ± 0.0003	0.3977 ± 0.0008	0.3966 ± 0	0.39 ± 0.0033	**0.4127 ± 0**	0.4094 ± 0
DGI↑	0.1156 ± 0.0196	0.1216 ± 0.0277	0.0891 ± 0.0120	0.1038 ± 0.0213	0.1064 ± 0.0001	0.1276 ± 0.0052	**0.2607 ± 0**	0.0890 ± 0
RLI↑	0.2781 ± 0.0013	0.2789 ± 0.0007	0.2793 ± 0	0.2792 ± 0.0001	**0.2795 ± 0**	0.2703 ± 0.0019	0.2764 ± 0	0.2793 ± 0
CHI↑	7.1011 ± 0.3862	7.2661 ± 0.2768	7.4209 ± 0.0017	7.4297 ± 0.0108	7.3349 ± 0.0002	5.0935 ± 0.201	6.5713 ± 0	**7.4499 ± 0**
RTI↓	1.0790 ± 0.5693	0.7392 ± 0.0896	0.7736 ± 0.0396	0.6885 ± 0.1221	0.6419 ± 0.0002	2.2234 ± 0.1331	0.8343 ± 0	**0.6145 ± 0**
WGI↑	0.4007 ± 0.021	0.4107 ± 0.0101	0.4077 ± 0.0003	0.4078 ± 0.0007	0.4009 ± 0.0001	0.3117 ± 0.0067	0.3707 ± 0	**0.4122 ± 0**
DI↑	0.0182 ± 0.0045	0.0218 ± 0.0047	0.0172 ± 0.0019	0.0177 ± 0.0017	0.0190 ± 0	0.0161 ± 0.0013	**0.0464 ± 0**	0.0156 ± 0
BHI↑	0.0033 ± 0.0002	0.0035 ± 0.0002	0.0036 ± 0.0001	0.0035 ± 0.0001	0.0035 ± 0	0.0033 ± 0.0001	**0.0038 ± 0**	0.0035 ± 0
PBMI↑	0.0140 ± 0.0016	0.0164 ± 0.0017	0.0180 ± 0.0003	0.0165 ± 0.0019	0.0171 ± 0	0.0096 ± 0	**0.0182 ± 0**	0.0157 ± 0
XBI↓	57.9946 ± 20.2405	50.3353 ± 23.5672	76.5522 ± 14.4527	64.5991 ± 19.5411	51.7592 ± 0.0010	67.5394 ± 9.8879	**12.0031 ± 0**	83.1830 ± 0
DBI↓	1.2280 ± 0.0617	1.1969 ± 0.0305	1.1965 ± 0.0083	1.1947 ± 0.0073	1.1875 ± 0.0003	1.4900 ± 0.0489	1.2158 ± 0	**1.1767 ± 0**
LSSRI↑	1.9588 ± 0.0559	1.9825 ± 0.0393	2.0043 ± 0.0002	2.0055 ± 0.0015	1.9926 ± 0	1.6271 ± 0.0430	1.8827 ± 0	**2.0082 ± 0**
TWI↓	1.4801 ± 0.074	1.4489 ± 0.0512	1.4206 ± 0.0003	1.4192 ± 0.0018	1.4353 ± 0	1.9656 ± 0.0747	1.5801 ± 0	**1.4158 ± 0**
ACC↑	0.5854 ± 0.0683	0.598 ± 0.046	0.6104 ± 0.0109	0.6035 ± 0.0132	0.5636 ± 0.0002	**0.6472 ± 0.0017**	0.6258 ± 0	0.5711 ± 0
NMI↑	0.6327 ± 0.0268	0.6345 ± 0.0161	0.6296 ± 0.0002	0.6319 ± 0.0029	0.6063 ± 0.0002	**0.7588 ± 0.0116**	0.6666 ± 0	0.6317 ± 0
ARI↑	0.4715 ± 0.052	0.4784 ± 0.0339	0.4785 ± 0.0034	0.4761 ± 0.0046	0.4469 ± 0.0001	**0.523 ± 0.0162**	0.4884 ± 0	0.4604 ± 0

neurodynamic optimization, is proposed for global optimization [59–62] by employing multiple modules or models (e.g., neural networks) for scattered searches coordinated by using a meta-heuristic rule for repositioning initial solutions. Integrating the scatter search capability of multiple neurodynamic models together with the global search capability of meta-heuristics, the collaborative neurodynamic optimization approach is proved to be almost surely convergent to global optima of optimization problems [59–63], and extended for multi-objective optimization [64,65], and combinatorial and mixed-integer optimization [62,63].

Several collaborative neurodynamic optimization paradigms have been developed for various applications such as model predictive control [66], nonnegative matrix factorization [67,68], vehicle-task assignment [69], stock portfolio selection [70], spiking neural network regularization [71], hash bit selection [72], and balanced clustering [29].

3. CAPKM++

The collaborative annealing power k-means++ clustering (CAPKM++) algorithm starts with initial cluster centers selected using k-means++, generates N sets of alternative centers using Power k-means, and updates the centers using a meta-heuristic rule for the subsequential regeneration of the centers repeatedly using Power k-means until no more improvement can be made.

As in most existing collaborative neurodynamic optimization paradigms, CAPKM++ uses the particle swarm optimization (PSO) rule defined in [73] as follows:

$$v_i(t+1) = c_0 v_i(t) + c_1 r_1(p_i^* - p_i(t)) + c_2 r_2(p^* - p_i(t)), \tag{11a}$$

$$p_i(t+1) = p_i(t) + v_i(t+1), \tag{11b}$$

where $p_i(t)$ is the position of the ith particle at the tth iteration, $v_i(t)$ is the velocity of the ith particle at the tth iteration, $p_i^*(t)$ is the present best position in terms of a given objective function for the ith particle, $p^*(t)$ is the best position of the population, c_0 is a positive inertia parameter, c_1, c_2 are two positive acceleration constants, and $r_1, r_2 \in [0, 1]$ are two random numbers.

The clustering performance of CAPKM++ depends heavily on the diversity of generated cluster centers. To enhance the diversity of alternative clusters, annealing parameters (i.e., $s(0)$ and η) in CAPKM++ are randomly generated in given ranges. In addition, a mutation operation is used in CAPKM++ to ensure a certain level of diversity. Specifically, if the diversity measure of the cluster centers matrix of different modules is low, a mutation operation is performed. A diversity measure is defined by:

$$\delta(\Theta) = \frac{1}{Nn}\sum_{j=1}^{N} \|\Theta^{(j)} - \Theta^*\|_2, \tag{12}$$

where Θ^* is the best cluster centers matrix among the N modules.

A wavelet mutation operator is used to diversify cluster centers to maintain the diversity of the cluster centers matrix. If the diversity is lower than a threshold (i.e., $\delta(\Theta) < \varepsilon$), the following wavelet mutation operation is performed:

$$\theta_i(t+1) = \begin{cases} \theta_i(t) + \tau(\overline{\theta}_i - \theta_i(t)) & \tau > 0, \\ \theta_i(t) + \tau(\theta_i(t) - \underline{\theta}_i) & \tau < 0, \end{cases} \tag{13}$$

where $\overline{\theta}_i$ and $\underline{\theta}_i$ are an upper bound and lower bound of the searching space of ith particle, and τ is defined by a wavelet function:

$$\tau = \frac{1}{\sqrt{a}} \exp -\frac{1}{2}(\frac{\psi}{a})^2 \cos(\frac{5\psi}{a}),$$

where ψ is randomly generated form interval $[-2.5a, 2.5a]$, $a = \exp 10(\ell/\ell_{\max})$, and $\ell_{\max}$ is the maximum iteration.

There are two hyperparameters (i.e., N and M) in CAPKM++, where N is the number of alternative centers generated and M is a termination criterion (i.e., if the best centers do not change for M times consecutively, the algorithm stops). These hyperparameters depend on the geometric shape and dimensionality of the given dataset. Usually, the more complex the data distribution is, or the larger the dataset is, the larger N and M need to be used.

Algorithm 1: CAPKM++

Input: Population size N, initial centers , velocity vector $v^{(i)} \in [-1, 1]^{mk}$, termination criterion M, termination counter l, penalty parameter ρ, particle/group best centers $\Theta^{(i)}/\ \Theta^*$, $f(\Theta^{(i)}, w^{(i)}) = f(\Theta^*, w^*) = \infty$, lower bound of initial power $\underline{s(0)}$, upper/lower bound of decrease rate $\underline{\eta}/\overline{\eta}$, PSO-based parameters c_0, c_1 and c_2, input data $X \in \mathbb{R}^{n\times m}$.

Output: w^*.

1 Select the initial centers $[\Theta_0^{(1)}, ..., \Theta_0^{(N)}] \in \{0, 1\}^{mk\times N}$ according to Eq. (3) **while** $l \le M$ **do**
2 **for** $i = 1$ **to** N **do**
3 Generate $s(0)$ and η randomly in the range of $[\underline{s}, 0)$ and $[\underline{\eta}, \overline{\eta}]$, respectively;
4 **repeat**
5 Update weights w according to Eq. (8);
6 Update centers Θ according to Eq. (9);
7 Reduce the power parameter s according to Eq. (10);
8 **until** *centers* Θ *converge*;
9 $\bar{\Theta}^{(i)} \leftarrow \Theta$;
10 $\bar{w}^{(i)} \leftarrow w$;
11 **if** $f(\bar{\Theta}^{(i)}, \bar{w}^{(i)}) < f(\Theta^{(i)}, w^{(i)})$ **then**
12 $\Theta^{(i)} \leftarrow \bar{\Theta}^{(i)}$;
13 $w^{(i)} \leftarrow \bar{w}^{(i)}$;
14 **end**
15 **end**
16 $i^* = \arg\min_i\{f(\Theta^{(1)}), ..., f(\Theta^{(i)}), ..., f(\Theta^{(N)}\}$;
17 **if** $f(\Theta^{(i^*)}, w^{(i^*)}) < f(\Theta^*, w^*)$ **then**
18 $\Theta^* \leftarrow \Theta^{(i^*)}$;
19 $w^* \leftarrow w^{(i^*)}$;
20 $l \leftarrow 0$;
21 **else**
22 $l \leftarrow l + 1$;
23 **end**
24 **for** $i = 1$ **to** N **do**
25 Update velocity according to Eq. (11a);
26 Update centers according to Eq. (11b);
27 **end**
28 Compute diversity $\delta(x)$ according to Eq. (12);
29 **if** $\delta(q) < \varepsilon$ **then**
30 Perform mutation according to Eq. (13);
31 **end**
32 **end**
33 **return** w^*.

Fig. 1 illustrates a flow chart of the CAPKM++ algorithm. Algorithm 1 details the pseudo-codes of the CAPKM++ algorithm. In steps 2–8, multiple power k-means clustering modules are employed to generate alternative clustering results. In step 3, an

Table 8
The mean values and standard deviations of internal and external cluster validity indices resulting from CAPKM++ and seven baselines on Optdigits and Satimage

Optdigits	KM	KM++	PKM	PKM++	EWPKM	SC	HC	CAPKM++
WGSS↓	236.9533 ± 1.6034	237.318 ± 2.583	235.6059 ± 1.1878	235.0509 ± 0.1682	395.162 ± 0.9684	246.8457 ± 0.3274	243.8351 ± 0	**234.8252 ± 0**
MRI↓	0.7171 ± 0.0055	0.7172 ± 0.0081	0.7126 ± 0.0039	**0.7107 ± 0.0003**	0.9433 ± 0.0016	0.7286 ± 0.0016	0.7253 ± 0	0.7111 ± 0
GPI↓	0.0181 ± 0.0012	0.0183 ± 0.0022	0.0172 ± 0.0009	**0.0168 ± 0**	0.1924 ± 0.0049	0.0233 ± 0.0005	0.0218 ± 0	0.0168 ± 0
BHGI↑	0.8204 ± 0.0121	0.8163 ± 0.0192	0.8240 ± 0.0084	0.8241 ± 0.0050	0.1982 ± 0.0067	0.7781 ± 0.0036	0.7862 ± 0	**0.8307 ± 0**
CI↓	0.1129 ± 0.0075	0.1154 ± 0.0108	0.1108 ± 0.0054	0.1114 ± 0.0041	0.3788 ± 0.0050	0.1255 ± 0.0015	0.1245 ± 0	**0.1059 ± 0**
TI↑	0.3689 ± 0.0145	0.3643 ± 0.0138	0.3644 ± 0.0114	0.3601 ± 0.0075	0.1373 ± 0.0036	0.3567 ± 0.0004	0.3552 ± 0	**0.3701 ± 0**
DGI↑	0.492 ± 0.0323	0.5085 ± 0.0367	0.4606 ± 0.0414	0.4526 ± 0.0304	0.3662 ± 0.0458	**0.6653 ± 0.0106**	0.6114 ± 0	0.4135 ± 0
RLI↑	0.1819 ± 0.0013	0.1814 ± 0.0016	0.1820 ± 0.0007	0.182 ± 0.0001	0.128 ± 0.0012	0.1784 ± 0.0001	0.1791 ± 0	**0.1821 ± 0**
CHI↑	0.8166 ± 0.0122	0.8139 ± 0.0194	0.8270 ± 0.0091	0.8312 ± 0.0013	0.0893 ± 0.0027	0.7437 ± 0.0023	0.7653 ± 0	**0.8330 ± 0**
RTI↓	1.4678 ± 0.5448	1.5858 ± 0.5395	1.4152 ± 0.3272	1.5324 ± 0.2959	18.2051 ± 1.7927	**1.0352 ± 0.0013**	1.0956 ± 0	1.1414 ± 0
WGI↑	0.2803 ± 0.0077	0.2769 ± 0.0096	0.2837 ± 0.0058	0.2829 ± 0.0057	0.0090 ± 0.0025	0.2739 ± 0.0004	0.2671 ± 0	**0.2904 ± 0**
DI↑	0.141 ± 0.0131	0.145 ± 0.0128	0.1286 ± 0.0145	0.1251 ± 0.0085	0.1101 ± 0.0137	**0.1741 ± 0.0026**	0.1692 ± 0	0.1139 ± 0
BHI↑	0.0416 ± 0.0004	0.0418 ± 0.0004	0.0417 ± 0.0002	0.0418 ± 0.0001	**0.0573 ± 0.0006**	0.042 ± 0.0001	0.0426 ± 0	0.0417 ± 0
PBMI↑	0.0025 ± 0.0001	0.0025 ± 0.0001	0.0025 ± 0.0001	0.0025 ± 0	0.0017 ± 0.0001	**0.0031 ± 0**	0.0026 ± 0	0.0025 ± 0
XBI↓	7.4884 ± 1.1402	6.9387 ± 0.978	8.7396 ± 1.6241	8.9998 ± 1.2776	15.4608 ± 4.1079	**4.0564 ± 0.1156**	4.3805 ± 0	10.6967 ± 0
DBI↓	1.8833 ± 0.0823	1.9158 ± 0.0877	1.8916 ± 0.0598	1.9111 ± 0.0423	4.9698 ± 0.2531	**1.8365 ± 0.0051**	1.9093 ± 0	1.8541 ± 0
LSSRI↑	−0.2027 ± 0.0151	−0.2061 ± 0.0243	−0.1901 ± 0.0111	−0.1848 ± 0.0016	−2.4166 ± 0.0301	−0.2961 ± 0.0031	−0.2675 ± 0	**−0.1827 ± 0**
TWI↓	23.6953 ± 0.1603	23.7318 ± 0.2583	23.5606 ± 0.1188	23.5051 ± 0.0168	39.5162 ± 0.0968	24.6846 ± 0.0327	24.3835 ± 0	**23.4825 ± 0**
ACC↑	0.7513 ± 0.0597	0.7506 ± 0.0512	0.7856 ± 0.0413	0.7988 ± 0.0049	0.1738 ± 0.004	**0.8264 ± 0.0003**	0.8089 ± 0	0.7913 ± 0
NMI↑	0.7304 ± 0.0271	0.7273 ± 0.0229	0.7468 ± 0.0188	0.7509 ± 0.0048	0.1358 ± 0.001	**0.8709 ± 0.0006**	0.8250 ± 0	0.7560 ± 0
ARI↑	0.6259 ± 0.061	0.6281 ± 0.0508	0.6621 ± 0.0427	0.6764 ± 0.0036	0.0151 ± 0.0012	**0.7797 ± 0.0003**	0.7170 ± 0	0.6702 ± 0
Satimage	**KM**	**KM++**	**PKM**	**PKM++**	**EWPKM**	**SC**	**HC**	**CAPKM++**
WGSS↓	44.353 ± 0.8238	44.6675 ± 1.0231	44.4344 ± 0.8616	44.2245 ± 0.6462	91.7690 ± 3.8351	49.1358 ± 0	51.3763 ± 0	**44.0144 ± 0**
MRI↓	0.3844 ± 0.0041	0.3863 ± 0.0056	0.3853 ± 0.0051	0.3841 ± 0.0038	0.5708 ± 0.0090	0.421 ± 0	0.4071 ± 0	**0.3829 ± 0**
GPI↓	0.0145 ± 0.0007	0.0148 ± 0.0009	0.0146 ± 0.0009	0.0144 ± 0.0006	0.0566 ± 0.0009	0.0236 ± 0	0.0201 ± 0	**0.0142 ± 0**
BHGI↑	0.893 ± 0.0016	0.8921 ± 0.0026	0.8923 ± 0.0025	0.8929 ± 0.0019	0.5767 ± 0.0148	0.8303 ± 0	0.8541 ± 0	**0.8934 ± 0**
CI↓	0.0441 ± 0.0007	0.0445 ± 0.001	0.0444 ± 0.001	0.0441 ± 0.0008	0.1726 ± 0.0076	0.0682 ± 0	0.0589 ± 0	**0.0439 ± 0**
TI↑	0.4644 ± 0.0063	**0.4667 ± 0.0078**	0.4651 ± 0.0065	0.4635 ± 0.0049	0.2986 ± 0.0156	0.4376 ± 0	0.4481 ± 0	0.4617 ± 0
DGI↑	0.34 ± 0.0429	0.3305 ± 0.0522	0.3359 ± 0.0369	0.3449 ± 0.0277	0.2391 ± 0.0319	0.3395 ± 0	0.3518 ± 0	**0.3539 ± 0**
RLI↑	0.3385 ± 0.0004	0.3384 ± 0.0004	0.3386 ± 0.0003	0.3386 ± 0.0002	0.3046 ± 0.0031	0.3343 ± 0	0.3310 ± 0	**0.3387 ± 0**
CHI↑	4.2644 ± 0.0944	4.2283 ± 0.1174	4.2549 ± 0.0991	4.2791 ± 0.0743	1.5479 ± 0.1104	3.7505 ± 0	3.5433 ± 0	**4.3032 ± 0**
RTI↓	0.4456 ± 0.0271	**0.4385 ± 0.0317**	0.4405 ± 0.0237	0.4463 ± 0.0178	3.6837 ± 0.7308	0.5033 ± 0	0.605 ± 0	0.4536 ± 0
WGI↑	0.481 ± 0.0034	0.4789 ± 0.0057	0.4796 ± 0.0058	0.481 ± 0.0044	0.1162 ± 0.0067	0.4515 ± 0	0.4019 ± 0	**0.4823 ± 0**
DI↑	0.0633 ± 0.0024	0.0632 ± 0.0034	0.0626 ± 0.0018	0.063 ± 0.0013	0.0373 ± 0.0016	0.0603 ± 0	0.055 ± 0	**0.0635 ± 0**
BHI↑	0.0077 ± 0.0005	0.0078 ± 0.0005	0.0076 ± 0.0003	0.0076 ± 0.0002	**0.0133 ± 0.0009**	0.0076 ± 0	0.0086 ± 0	0.0075 ± 0
PBMI↑	0.0256 ± 0.0004	0.0257 ± 0.0005	**0.0258 ± 0.0002**	0.0258 ± 0.0002	0.0110 ± 0.0009	0.0245 ± 0	0.0219 ± 0	0.0257 ± 0
XBI↓	11.1579 ± 1.1594	11.3685 ± 1.6075	11.4392 ± 1.0936	11.1727 ± 0.8202	27.6036 ± 2.9885	12.1163 ± 0	**9.5751 ± 0**	10.9062 ± 0
DBI↓	1.0299 ± 0.0601	1.0527 ± 0.0765	1.0343 ± 0.064	1.0187 ± 0.048	2.4885 ± 0.0928	1.026 ± 0	1.1769 ± 0	**1.0041 ± 0**
LSSRI↑	1.4501 ± 0.0226	1.4414 ± 0.0281	1.4478 ± 0.0237	1.4536 ± 0.0178	0.4346 ± 0.0693	1.3219 ± 0	1.2651 ± 0	**1.4594 ± 0**
TWI↓	6.3361 ± 0.1177	6.3811 ± 0.1462	6.3478 ± 0.1231	6.3178 ± 0.0923	13.1099 ± 0.5479	7.0194 ± 0	7.3395 ± 0	**6.2878 ± 0**
ACC↑	0.7048 ± 0.0267	0.6924 ± 0.0370	0.6984 ± 0.0349	0.7069 ± 0.0262	0.4557 ± 0.0089	0.6855 ± 0	0.6828 ± 0	**0.7152 ± 0**
NMI↑	0.6292 ± 0.007	0.6259 ± 0.0102	0.6276 ± 0.0100	0.6301 ± 0.0075	0.3729 ± 0.0042	**0.6340 ± 0**	0.6041 ± 0	0.6324 ± 0
ARI↑	0.5621 ± 0.0172	0.5543 ± 0.0238	0.5588 ± 0.0226	0.5643 ± 0.0169	0.2452 ± 0.0025	0.5032 ± 0	0.5488 ± 0	**0.5694 ± 0**

initial exponent $s(0)$ and discount factor η are randomly generated in given ranges. In steps 11–14, the best set of centers is determined via winner-takes-all. In steps 16–23, the best matrix of cluster centers is determined. In steps 24–27, the centers are re-initialized using the PSO rule. In step 28, the diversity of the centers among alternative solutions is measured. In steps 29–31, a wavelet mutation operation is carried out, where ε in step 22 is a threshold.

4. Experiments

4.1. Performance criteria

Nineteen internal cluster validity indices in Table 3 and three external indices are used to evaluate the clustering performance.

Accuracy (ACC) is defined as the ratio of the number of correctly assigned sample points to the total number of sample points, and it is stated as follows [74]:

$$ACC = \frac{\#\text{ of correct decisions}}{\#\text{ of total decisions}},$$

Let L denotes the set of cluster labels and $\hat{L}$ denotes the set of clusters obtained from a given clustering algorithm. Their mutual information is defined as follows:

$$MI(L,\hat{L}) = \sum_{l_i\in L,\hat{l}_j\in\hat{L}} p(l_i,\hat{l}_j)\log\frac{p(l_i,\hat{l}_j)}{p(l_i)p(\hat{l}_j)},$$

where $p(l_i)$ and $p(\hat{l}_j)$ are the probabilities that an arbitrary sample of the dataset belongs to the clusters l_i and $\hat{l}_j$, respectively. $p(l_i,\hat{l}_j)$ is the joint probability that the clusters l_i as well as $\hat{l}_j$. The probabilities are estimated from the given datasets.

Normalized mutual information (NMI) [74] is a normalization of the mutual information (MI) score to scale the results between 0 (no mutual information) and 1 (perfect correlation). It is defined as follows:

$$NMI(L,\hat{L}) = \frac{MI(L,\hat{L})}{\sqrt{H(L)\cdot H(\hat{L})}},$$

where $H(X) = -\sum_{x_i\in X} p(x_i)\log p(x_i)$ is the entropy of the clusters for X.

Table 9
The mean values and standard deviations of internal and external cluster validity indices resulting from CAPKM++ and seven baselines on COIL2000 and Penbased.

COIL2000	KM	KM++	PKM	PKM++	EWPKM	SC	HC	CAPKM++
WGSS↓	222.9596 ± 0.0053	222.9601 ± 0.0052	222.9554 ± 0	222.9554 ± 0	251.1241 ± 0	264.0148 ± 0.0146	229.2226 ± 0	**222.9522 ± 0**
MRI↓	0.8286 ± 0.0004	**0.8286 ± 0.0004**	0.829 ± 0	0.829 ± 0	0.948 ± 0	0.9995 ± 0.0003	0.8565 ± 0	0.8292 ± 0
GPI↓	**0.1225 ± 0.0001**	0.1225 ± 0.0001	0.1226 ± 0	0.1226 ± 0	0.2154 ± 0	0.2499 ± 0.0002	0.1443 ± 0	0.1225 ± 0
BHGI↑	0.5015 ± 0.0007	**0.5016 ± 0.0007**	0.5007 ± 0	0.5007 ± 0	0.1384 ± 0	0.0006 ± 0.0006	0.4137 ± 0	0.5003 ± 0
CI↓	0.2325 ± 0.0008	**0.2324 ± 0.0008**	0.2333 ± 0	0.2333 ± 0	0.4233 ± 0	0.4993 ± 0.0004	0.2805 ± 0	0.2337 ± 0
TI↑	0.3516 ± 0.0008	**0.3517 ± 0.0008**	0.3508 ± 0	0.3508 ± 0	0.0978 ± 0	0.0004 ± 0.0004	0.2902 ± 0	0.3504 ± 0
DGI↑	0.1614 ± 0.0692	0.153 ± 0.0618	0.2955 ± 0	0.2955 ± 0	0.4408 ± 0	0.4783 ± 0.0014	**0.5772 ± 0**	0.1286 ± 0
RLI↑	0.1863 ± 0.0004	**0.1863 ± 0.0004**	0.1861 ± 0	0.1861 ± 0	0.1081 ± 0	0.0113 ± 0.0026	0.1658 ± 0	0.1859 ± 0
CHI↑	0.1843 ± 0	0.1843 ± 0	0.1843 ± 0	0.1843 ± 0	0.0515 ± 0	0.0004 ± 0.0002	0.1519 ± 0	**0.1844 ± 0**
RTI↓	1.1797 ± 0.0090	1.1807 ± 0.0086	1.1743 ± 0	1.1743 ± 0	4.8548 ± 0	967.3506 ± 660.5126	1.4395 ± 0	**1.1693 ± 0**
WGI↑	0.2685 ± 0.0017	0.2683 ± 0.0016	0.2696 ± 0	0.2696 ± 0	0.1035 ± 0	0.0009 ± 0.0005	0.2384 ± 0	**0.2704 ± 0**
DI↑	0.0347 ± 0.0147	0.0329 ± 0.0131	0.0633 ± 0	0.0633 ± 0	0.0982 ± 0	0.1087 ± 0.001	**0.121 ± 0**	0.0275 ± 0
BHI↑	0.0229 ± 0.0001	0.0229 ± 0	0.0229 ± 0	0.0229 ± 0	0.0256 ± 0	**0.0269 ± 0**	0.0233 ± 0	0.0228 ± 0
PBMI↑	0.0057 ± 0	0.0057 ± 0	0.0057 ± 0	0.0057 ± 0	0.0014 ± 0	0 ± 0	0.0047 ± 0	**0.0058 ± 0**
XBI↓	103.4457 ± 41.1282	108.4566 ± 36.7143	23.2723 ± 0	23.2723 ± 0	10.6284 ± 0	9.0988 ± 0.0213	**6.2858 ± 0**	123.484 ± 0
DBI↓	2.1258 ± 0.0106	2.1269 ± 0.0100	2.1192 ± 0	2.1192 ± 0	4.2882 ± 0	58.0145 ± 18.2979	2.3327 ± 0	**2.1134 ± 0**
LSSRI↑	−1.6912 ± 0.0002	−1.6912 ± 0.0001	−1.6910 ± 0	−1.6910 ± 0	−2.9666 ± 0	−8.0848 ± 0.5934	−1.8842 ± 0	**−1.6911 ± 0**
TWI↓	111.4798 ± 0.0027	111.4801 ± 0.0026	111.4777 ± 0	111.4777 ± 0	125.562 ± 0	131.9784 ± 0.0267	114.6113 ± 0	**111.4761 ± 0**
ACC↑	0.6797 ± 0.004	0.6792 ± 0.0038	0.6821 ± 0	0.6821 ± 0	0.5228 ± 0	0.504 ± 0.0033	0.6742 ± 0	**0.6842 ± 0**
NMI↑	0.011 ± 0.0001	0.011 ± 0.0001	0.0111 ± 0	0.0111 ± 0	**0.0221 ± 0**	0.0002 ± 0.0002	0.0087 ± 0	0.0111 ± 0
ARI↑	0.0313 ± 0.0010	0.0312 ± 0.0009	0.0319 ± 0	0.0319 ± 0	0.0018 ± 0	0 ± 0.0001	0.027 ± 0	**0.0322 ± 0**
Penbased	**KM**	**KM++**	**PKM**	**PKM++**	**EWPKM**	**SC**	**HC**	**CAPKM++**
WGSS↓	317.5451 ± 5.4898	318.28 ± 7.002	317.3646 ± 4.125	319.1628 ± 3.8478	1021.8032 ± 0	396.8291 ± 0	322.8555 ± 0	**308.1344 ± 0**
MRI↓	0.5083 ± 0.0078	0.5114 ± 0.0132	0.5036 ± 0.0025	0.5045 ± 0.0031	0.5932 ± 0	0.5932 ± 0	0.5122 ± 0	**0.4995 ± 0**
GPI↓	0.0096 ± 0.0011	0.0102 ± 0.0018	0.0098 ± 0.0005	0.0100 ± 0.0005	0.0274 ± 0	0.0274 ± 0	0.0101 ± 0	**0.0078 ± 0**
BHGI↑	0.908 ± 0.0131	0.902 ± 0.0146	0.8962 ± 0.0059	0.8939 ± 0.0058	0.8028 ± 0	0.8028 ± 0	0.9099 ± 0	**0.9282 ± 0**
CI↓	0.0723 ± 0.0113	0.0773 ± 0.0108	0.0829 ± 0.0035	0.0845 ± 0.0032	0.1179 ± 0	0.1179 ± 0	0.0654 ± 0	**0.0565 ± 0**
TI↑	0.4157 ± 0.0187	0.4109 ± 0.0172	0.3900 ± 0.0050	0.3881 ± 0.0044	0.4234 ± 0	0.4234 ± 0	0.431 ± 0	**0.4335 ± 0**
DGI↑	0.1536 ± 0.0048	0.1508 ± 0.0086	0.1533 ± 0.0071	0.1505 ± 0.0056	0.3336 ± 0	**0.3336 ± 0**	0.2361 ± 0	0.1542 ± 0
RLI↑	0.2545 ± 0.0012	0.2543 ± 0.0016	0.2545 ± 0.0008	0.2543 ± 0.0008	0.2381 ± 0	0.2381 ± 0	0.2539 ± 0	**0.2562 ± 0**
CHI↑	2.2187 ± 0.0554	2.2118 ± 0.0693	2.2202 ± 0.0417	2.202 ± 0.0388	1.5749 ± 0	1.5749 ± 0	2.1649 ± 0	**2.3161 ± 0**
RTI↓	0.8736 ± 0.1768	0.9673 ± 0.1305	0.9858 ± 0.0437	0.9699 ± 0.0423	1.328 ± 0	1.328 ± 0	**0.583 ± 0**	0.6533 ± 0
WGI↑	0.4182 ± 0.0185	0.4153 ± 0.0124	0.4028 ± 0.0066	0.4014 ± 0.0056	0.3582 ± 0	0.3582 ± 0	0.4201 ± 0	**0.4383 ± 0**
DI↑	0.0336 ± 0.0016	0.0330 ± 0.0016	0.0336 ± 0.0019	0.0333 ± 0.0018	0.0761 ± 0	**0.0761 ± 0**	0.0523 ± 0	0.0358 ± 0
BHI↑	0.029 ± 0.0008	0.0289 ± 0.0009	0.0295 ± 0.0005	0.0297 ± 0.0005	0.0297 ± 0	**0.0297 ± 0**	0.0289 ± 0	0.0283 ± 0
PBMI↑	0.0108 ± 0.0008	0.0106 ± 0.0007	0.0107 ± 0.0006	0.0106 ± 0.0006	0.0113 ± 0	0.0113 ± 0	**0.0121 ± 0**	0.012 ± 0
XBI↓	69.3926 ± 4.1607	71.3386 ± 5.8851	66.9503 ± 0.8702	67.3297 ± 0.8117	**10.8986 ± 0**	10.8986 ± 0	26.5508 ± 0	65.0032 ± 0
DBI↓	1.3362 ± 0.0711	1.3461 ± 0.0569	1.4207 ± 0.0339	1.4361 ± 0.0117	1.4269 ± 0	1.4269 ± 0	**1.2659 ± 0**	1.2696 ± 0
LSSRI↑	0.7966 ± 0.0251	0.7933 ± 0.0318	0.7974 ± 0.0188	0.7892 ± 0.0176	0.4542 ± 0	0.4542 ± 0	0.7724 ± 0	**0.8399 ± 0**
TWI↓	31.7545 ± 0.549	31.828 ± 0.7002	31.7365 ± 0.4125	31.9163 ± 0.3848	39.6829 ± 0	39.6829 ± 0	32.2856 ± 0	**30.8134 ± 0**
ACC↑	0.6853 ± 0.0423	0.6939 ± 0.0440	**0.7431 ± 0.0154**	0.7364 ± 0.0156	0.1041 ± 0	0.7248 ± 0	0.682 ± 0	0.6671 ± 0
NMI↑	0.6802 ± 0.0097	0.6817 ± 0.0144	0.6947 ± 0.0107	0.6968 ± 0.0116	0 ± 0	**0.7851 ± 0**	0.7284 ± 0	0.6821 ± 0
ARI↑	0.5468 ± 0.0320	0.5541 ± 0.0396	0.5981 ± 0.0228	**0.6016 ± 0.0254**	0 ± 0	0.5665 ± 0	0.5532 ± 0	0.5318 ± 0

The adjusted rand index (ARI) [75] is defined as follows:

$$ARI(L,\hat{L}) = \frac{\sum_{i,j}\binom{n_{ij}}{2} - \left[\sum_i \binom{n_i}{2}\sum_j \binom{\hat{n}_j}{2}\right]/\binom{n}{2}}{\frac{1}{2}\left[\sum_i \binom{n_i}{2} + \sum_j \binom{\hat{n}_j}{2}\right] - \left[\sum_i \binom{n_i}{2}\sum_j \binom{\hat{n}_j}{2}\right]/\binom{n}{2}},$$

where $n_{i,j}$ is the number of points that are in cluster L_i and cluster $\hat{L}_j$, n_i is the number of points that are in cluster L_i, $\hat{n}_j$ is the number of points that are in cluster $\hat{L}_j$, and $\binom{n}{m} = n!/(m!(n-m)!)$ is the number of combinations of n objects taken m at a time.

4.2. Setups

Twelve benchmark datasets are used in the experiments: NCI9,[1] Lymphoma,[2] ORL10P,[3] WarpPIE10P,[4] Segment,[5] SpamBase,[6] PageBlocks,[7] Texture,[8] Optdigits,[9] Satimage,[10] COIL2000,[11] and Penbased.[12] Table 1 lists their major information.

Seven baselines (i.e., k-means (KM),[13] k-mean++ (KM++),[14] Power k-means (PKM) with its codes shared by the first author of [48], Power k-means starting with k-mean++ (PKM++), entropy weighted power k-means (EWPKM),[15] spectral clustering (SC),[16] hierarchical clustering (HC)[17] are used for performance comparison. The hierarchical clustering algorithm used is an agglomerative one called the minimum variance algorithm in MATLAB. The

1 https://jundongl.github.io/scikit-feature/files/datasets/nci9.mat
2 https://jundongl.github.io/scikit-feature/files/datasets/lymphoma.mat
3 https://jundongl.github.io/scikit-feature/files/datasets/orlraws10P.mat
4 https://jundongl.github.io/scikit-feature/files/datasets/warpPIE10P.mat
5 https://sci2s.ugr.es/keel/dataset.php?cod=107
6 https://sci2s.ugr.es/keel/dataset.php?cod=109
7 https://sci2s.ugr.es/keel/dataset.php?cod=104
8 https://sci2s.ugr.es/keel/dataset.php?cod=72
9 https://sci2s.ugr.es/keel/dataset.php?cod=199
10 https://sci2s.ugr.es/keel/dataset.php?cod=71
11 https://sci2s.ugr.es/keel/dataset.php?cod=164
12 https://sci2s.ugr.es/keel/dataset.php?cod=70
13 https://www.mathworks.com/help/stats/kmeans.html?s_tid=srchtitle_kmean_1
14 https://github.com/xuyxu/Clustering
15 https://github.com/DebolinaPaul/EWP
16 https://www.mathworks.com/help/stats/spectralcluster.html
17 https://www.mathworks.com/help/stats/hierarchical-clustering.html?s_tid=srchtitle_hierarchical%20clustering_1

codes of CAPKM++ are publicly available from https://github.com/HongzongLI-CS/CPKM-Github. The Euclidean distance is used as the dissimilarity metric in all algorithms.

In this study, The hyperparameters values (i.e., N and M) of CAPKM++ are determined based on the Monte Carlo test results. Fig. 5 illustrates the Monte Carlo test results obtained by the CAPKM++ algorithm on the twelve benchmark datasets. Table 1 lists the values of the hyper-parameters used in the experiments, where the objective function values reach their minimum in most runs with the listed hyper-parameters value. In PSO rule (11), $c_0 = c_1 = c_2 = 1$. The diversity threshold ε is set to be 0.004. In the power k-means module, the exponent parameter $s(0)$ and the decreasing parameter η are randomly generated at the range of $[-20, 0)$ and $[1.1, 1.4]$, respectively.

4.3. Ablation studies

Table 2 records the mean values and standard deviations of $f(\Theta)$ in (1), using PKM, CAPKM++, and two intermediate algorithms with add-on parts, over 50 runs with random initialization on the twelve datasets, where PKM++ denotes Power k-means started with k-means++, CAPKM denotes collaborative annealing power k-means with multiple Power k-means modules re-initialized using PSO, and the best and second-best results are boldfaced and underlined, respectively. As shown in Table 2, PKM++ achieves better performance than PKM statistically in terms of $f(\Theta)$ mean value on 10 out of the 12 datasets (i.e., 83.3%), and CAPKM++ achieves the best performance in terms of mean value on all datasets (i.e., 100%).

4.4. Experimental results

Figs. 2–3 depict twelve snapshots of the convergent weights w in eqn. (8) and the descending objective function $f(\Theta)$ in eqn. (1) using PKM in the inner-loop of CAPKM++ (Steps 4–8) on the twelve benchmark datasets. The snapshots in Fig. 2 show that the weights converge within 200 iterations and those in Fig. 3 show that the objective function values monotonically decrease and reach their minima within 150 iterations in the inner-loops of CAPKM++. Fig. 4 depicts the convergent behaviors of $f(\Theta)$ the CAPKM++ algorithm on the twelve datasets, where the lower envelopes illustrate the objective function values of population-best solutions Θ^*. It shows that CAPKM++ converges within 80 iterations.

To visualize the clustering performance of the high-dimensional data, Figs. 6–7 display the clustering results of CAPKM++ and baselines on two representative datasets mapped to two-dimensional planes using the t-distributed stochastic neighborhood embedding (t-SNE) algorithm [76]. As shown in Figs. 6–7, CAPKM++ achieves better clustering results than the other algorithms.

Tables 4–9 summarize the mean values and standard deviations of twenty internal cluster validity indices and three external indices resulting from the seven competing algorithms over 50 runs with random initialization on the twelve datasets, where the best and second-best results are boldfaced and underlined, respectively. In view that five internal indices have large value ranges, to facilitate the tabular presentation, WGSS, CHI, BHI, PBMI, TWI value are normalized by m, $(n-k)/(k-1)$, m, m, mk, respectively. Specifically, CAPKM++ achieves the 145 best and 31 second-best mean values out of 252 entries (i.e., 57.54% and 69.84% in terms of best values and best plus second-best values) among the competing algorithms, whereas HC achieves the second-best performance with 33 best and 71 second-best means out of 252 entries (i.e., 13.10% and 40.87% in terms of best and best plus second-best mean values). In addition, the standard deviations of the results of CAPKM++ are zero, indicating the highest consistency of the results.

5. Concluding remarks

This paper presents a clustering algorithm based on k-means++, collaborative neurodynamic optimization, and power k-means. The proposed clustering algorithm incorporates k-means++ in initialization, leverages the clustering capability of Power k-means, and uses a particle swarm optimization rule to reposition the centers repeatedly upon their local convergence. The proposed algorithm statistically outperforms the baselines owing to the combined use of k-means++ to select better initial centers and Power k-means with random annealing parameters to generate more alternative cluster centers in the collaborative neurodynamic optimization framework. Further research may aim to the extension of the proposed algorithm for constrained clustering and its applications in the areas of logistics and finance.

CRediT authorship contribution statement

Hongzong Li: Data curation, Software, Investigation, Validation, Writing – original draft. **Jun Wang:** Conceptualization, Methodology, Writing – review & editing, Funding acquisition, Resources, Supervision, Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] R. Xu, D. Wunsch, Survey of clustering algorithms, IEEE Trans. Neural Netw. 16 (2005) 645–678.
[2] A.E. Ezugwu, A.M. Ikotun, O.O. Oyelade, L. Abualigah, J.O. Agushaka, C.I. Eke, A.A. Akinyelu, A comprehensive survey of clustering algorithms: State-of-the-art machine learning applications, taxonomy, challenges, and future research prospects, Eng. Appl. Artif. Intell. 110 (2022) 104743.
[3] A.K. Jain, M.N. Murty, P.J. Flynn, Data clustering: a review, ACM Comput. Surv. 31 (1999) 264–323.
[4] S.C. Johnson, Hierarchical clustering schemes, Psychometrika 32 (1967) 241–254.
[5] P. Macnaughton-Smith, W. Williams, M. Dale, L. Mockett, Dissimilarity analysis: a new technique of hierarchical sub-division, Nature 202 (1964) 1034–1035.
[6] S. Lloyd, Least squares quantization in PCM, IEEE Trans. Inform. Theory 28 (1982) 129–137.
[7] B. Zhang, Generalized k-harmonic means–dynamic weighting of data in unsupervised learning, in: Proceedings of the 2001 SIAM International Conference on Data Mining, SIAM, 2001, pp. 1–13.
[8] J. Wang, A linear assignment clustering algorithm based on the least similar cluster representatives, IEEE Trans. Syst. Man Cybern. A 29 (1999) 100–104.
[9] J. Shi, J. Malik, Normalized cuts and image segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 22 (2000) 888–905.
[10] A. Ng, M. Jordan, Y. Weiss, On spectral clustering: Analysis and an algorithm, in: T. Dietterich, S. Becker, Z. Ghahramani (Eds.), Advances in Neural Information Processing Systems, Volume 14, MIT Press, 2001.
[11] Q. Wang, Z. Qin, F. Nie, X. Li, Spectral embedded adaptive neighbors clustering, IEEE Trans. Neural Netw. Learn. Syst. 30 (2018) 1265–1271.
[12] J. Bilmes, A Gentle Tutorial of the EM Algorithm and Its Application To Parameter Estimation for Gaussian Mixture and Hidden Markov Models, Technical Report TR-97-021, International Computer Science Institute, 1997.
[13] W. Chen, X. Wang, Z. Cai, C. Liu, Y. Zhu, W. Lin, DP-GMM clustering-based ensemble learning prediction methodology for dam deformation considering spatiotemporal differentiation, Knowl.-Based Syst. 222 (2021) 106964.
[14] D. Comaniciu, P. Meer, Mean shift: A robust approach toward feature space analysis, IEEE Trans. Pattern Anal. Mach. Intell. 24 (2002) 603–619.
[15] W. Guo, W. Wang, S. Zhao, Y. Niu, Z. Zhang, X. Liu, Density peak clustering with connectivity estimation, Knowl.-Based Syst. 243 (2022) 108501.
[16] R. Krishnapuram, J.M. Keller, A possibilistic approach to clustering, IEEE Trans. Fuzzy Syst. 1 (1993) 98–110.

[17] M.J. Li, M.K. Ng, Y. Cheung, J.Z. Huang, Agglomerative fuzzy k-means clustering algorithm with selection of number of clusters, IEEE Trans. Knowl. Data Eng. 20 (2008) 1519–1534.
[18] Y. Lin, S. Chen, A centroid auto-fused hierarchical fuzzy c-means clustering, IEEE Trans. Fuzzy Syst. 29 (2020) 2006–2017.
[19] J. Zhou, W. Pedrycz, C. Gao, Z. Lai, J. Wan, Z. Ming, Robust jointly sparse fuzzy clustering with neighborhood structure preservation, IEEE Trans. Fuzzy Syst. 30 (2022) 1073–1087.
[20] Y. Gao, Z. Wang, J. Xie, J. Pan, A new robust fuzzy c-means clustering method based on adaptive elastic distance, Knowl.-Based Syst. 237 (2022) 107769.
[21] L. Jing, M.K. Ng, J.Z. Huang, An entropy weighting k-means algorithm for subspace clustering of high-dimensional sparse data, IEEE Trans. Knowl. Data Eng. 19 (2007) 1026–1041.
[22] S. Chakraborty, D. Paul, S. Das, J. Xu, Entropy weighted power k-means clustering, in: International Conference on Artificial Intelligence and Statistics, PMLR, 2020, pp. 691–701.
[23] S. Chakraborty, S. Das, Detecting meaningful clusters from high-dimensional data: A strongly consistent sparse center-based clustering approach, IEEE Trans. Pattern Anal. Mach. Intell. (2022) in press.
[24] T. Zhang, P. Ji, M. Harandi, W. Huang, H. Li, Neural collaborative subspace clustering, in: International Conference on Machine Learning, PMLR, 2019, pp. 7384–7393.
[25] Q. Zheng, J. Zhu, Z. Tian, Z. Li, S. Pang, X. Jia, Constrained bilinear factorization multi-view subspace clustering, Knowl.-Based Syst. 194 (2020) 105514.
[26] L. Wei, F. Zhang, Z. Chen, R. Zhou, C. Zhu, Subspace clustering via adaptive least square regression with smooth affinities, Knowl.-Based Syst. 239 (2022) 107950.
[27] K.S. Al-Sultan, A tabu search approach to the clustering problem, Pattern Recognit. 28 (1995) 1443–1451.
[28] S.Z. Selim, K. Alsultan, A simulated annealing algorithm for the clustering problem, Pattern Recognit. 24 (1991) 1003–1008.
[29] X. Dai, J. Wang, W. Zhang, Balanced clustering based on collaborative neurodynamic optimization, in: Kownlesge-Based Systems, Vol. 250, 2022.
[30] J. MacQueen, et al., Some methods for classification and analysis of multivariate observations, in: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Oakland, CA, USA volume 1, 1967, pp. 281–297.
[31] A.W. Edwards, L.L. Cavalli-Sforza, A method for cluster analysis, Biometrics 21 (1965) 362–375.
[32] G.H. Ball, D.J. Hall, ISODATA, a Novel Method of Data Analysis and Pattern Classification, Technical Report, Stanford Research Institute, Menlo Park, CA, USA, 1965.
[33] F.J. Rohlf, Methods of comparing classifications, Annu. Rev. Ecol. Syst. 5 (1974) 101–113.
[34] T. Caliński, J. Harabasz, A dendrite method for cluster analysis, Comm. Statist. Theory Methods 3 (1974) 1–27.
[35] J.C. Dunn, Well-separated clusters and optimal fuzzy partitions, J. Cybern. 4 (1974) 95–104.
[36] F.B. Baker, L.J. Hubert, Measuring the power of hierarchical cluster analysis, J. Amer. Statist. Assoc. 70 (1975) 31–38.
[37] J.O. McClain, V.R. Rao, Clustisz: A program to test for the quality of clustering of a set of objects, J. Mar. Res. 12 (1975) 456–460.
[38] J.A. Hartigan, Clustering Algorithms, John Wiley & Sons, Inc., USA, 1975.
[39] L. Hubert, J. Schultz, Quadratic assignment as a general data analysis strategy, Br. J. Math. Stat. Psychol. 29 (1976) 190–241.
[40] D. Ratkowsky, G. Lance, Criterion for determining the number of groups in a classification, Aust. Comput. J. 10 (1978) 115–117.
[41] D.L. Davies, D.W. Bouldin, A cluster separation measure, IEEE Trans. Pattern Anal. Mach. Intell. 1 (1979) 224–227.
[42] G.W. Milligan, A monte carlo study of thirty internal criterion measures for cluster analysis, Psychometrika 46 (1981) 187–199.
[43] X.L. Xie, G. Beni, A validity measure for fuzzy clustering, IEEE Trans. Pattern Anal. Mach. Intell. 13 (1991) 841–847.
[44] J.C. Bezdek, N.R. Pal, Some new indexes of cluster validity, IEEE Trans. Syst. Man Cybern. B 28 (1998) 301–315.
[45] S. Ray, R.H. Turi, Determination of number of clusters in k-means clustering and application in colour image segmentation, in: Proc. of 4th International Conference on Advances in Pattern Recognition and Digital Techniques, 1999, pp. 137–143.
[46] B. Desgraupes, Clustering indices, University of Paris Ouest-Lab Modal'X 1 (34) (2013).
[47] D. Arthur, S. Vassilvitskii, How slow is the k-means method? in: Proceedings of the 22nd Annual Symposium on Computational Geometry, 2006, pp. 144–153.
[48] J. Xu, K. Lange, Power k-means clustering, in: International Conference on Machine Learning, PMLR, 2019, pp. 6921–6931.
[49] P.S. Bradley, U.M. Fayyad, Refining initial points for k-means clustering, in: International Conference on Machine Learning, Volume 98, PMLR, 1998, pp. 91–99.
[50] D. Arthur, S. Vassilvitskii, K-Means++: The Advantages of Careful Seeding, Technical Report, Stanford University, 2006.
[51] O. Bachem, M. Lucic, H. Hassani, A. Krause, Fast and provably good seedings for k-means, in: D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, R. Garnett (Eds.), Advances in Neural Information Processing Systems, Volume 29, Curran Associates, Inc., 2016.
[52] H. Zha, X. He, C. Ding, M. Gu, H. Simon, Spectral relaxation for k-means clustering, in: T. Dietterich, S. Becker, Z. Ghahramani (Eds.), Advances in Neural Information Processing Systems, Volume 14, MIT Press, 2001.
[53] Z. Güngör, A. Ünler, K-harmonic means data clustering with simulated annealing heuristic, Appl. Math. Comput. 184 (2007) 199–209.
[54] F. Yang, T. Sun, C. Zhang, An efficient hybrid data clustering method based on k-harmonic means and particle swarm optimization, Expert Syst. Appl. 36 (2009) 9847–9852.
[55] O. Du Merle, P. Hansen, B. Jaumard, N. Mladenovic, An interior point algorithm for minimum sum-of-squares clustering, SIAM J. Sci. Comput. 21 (2000) 1485–1505.
[56] D. Aloise, P. Hansen, L. Liberti, An improved column generation algorithm for minimum sum-of-squares clustering, Math. Program. 131 (2012) 195–220.
[57] D. Aloise, A. Deshpande, P. Hansen, P. Popat, Np-hardness of Euclidean sum-of-squares clustering, Mach. Learn. 75 (2009) 245–248.
[58] D.W. Cantrell, E.W. Weisstein, Power Mean. From MathWorld–A Wolfram Web Resource, https://mathworld.wolfram.com/PowerMean.html.
[59] Z. Yan, J. Wang, G. Li, A collective neurodynamic optimization approach to bound-constrained nonconvex optimization, Neural Netw. 55 (2014) 20–29.
[60] Z. Yan, J. Fan, J. Wang, A collective neurodynamic approach to constrained global optimization, IEEE Trans. Neural Netw. Learn. Syst. 28 (2017) 1206–1215.
[61] H. Che, J. Wang, A two-timescale duplex neurodynamic approach to biconvex optimization, IEEE Trans. Neural Netw. Learn. Syst. 30 (2019b) 2503–2514.
[62] H. Che, J. Wang, A collaborative neurodynamic approach to global and combinatorial optimization, Neural Netw. 114 (2019a) 15–27.
[63] Che. H., J. Wang, A two-timescale duplex neurodynamic approach to mixed-integer optimization, IEEE Trans. Neural Netw. Learn. Syst. 32 (2021) 36–48.
[64] S. Yang, Q. Liu, J. Wang, A collaborative neurodynamic approach to multiple-objective distributed optimization, IEEE Trans. Neural Netw. Learn. Syst. 29 (2018) 981–992.
[65] M.-F. Leung, J. Wang, A collaborative neurodynamic approach to multi-objective optimization, IEEE Trans. Neural Netw. Learn. Syst. 29 (2018) 5738–5748.
[66] Z. Yan, J. Wang, Nonlinear model predictive control based on collective neurodynamic optimization, IEEE Trans. Neural Netw. Learn. Syst. 26 (2015) 840–850.
[67] J. Fan, J. Wang, A collective neurodynamic optimization approach to nonnegative matrix factorization, IEEE Trans. Neural Netw. Learn. Syst. 28 (2017) 2344–2356.
[68] H. Che, J. Wang, A. Cichocki, Bicriteria sparse nonnegative matrix factorization via two-timescale duplex neurodynamic optimization, IEEE Trans. Neural Netw. Learn. Syst. (2022) in press.
[69] J. Wang, J. Wang, Q.-L. Han, Multi-vehicle task assignment based on collaborative neurodynamic optimization with discrete Hopfield networks, IEEE Trans. Neural Netw. Learn. Syst. 32 (2021) 5274–5286.
[70] M.-F. Leung, J. Wang, Minimax and biobjective portfolio selection based on collaborative neurodynamic optimization, IEEE Trans. Neural Netw. Learn. Syst. 32 (2021) 2825–2836.
[71] J. Zhao, J. Yang, J. Wang, W. Wu, Spiking neural network regularization with fixed and adaptive drop-keep probabilities, IEEE Trans. Neural Netw. Learn. Syst. 33 (2022) 4096–4109.
[72] X. Li, J. Wang, S. Kwong, Hash bit selection via collaborative neurodynamic optimization with discrete Hopfield networks, IEEE Trans. Neural Netw. Learn. Syst. 32 (2022) in press.
[73] J. Kennedy, R. Eberhart, Particle swarm optimization, in: Proceedings of ICNN'95-International Conference on Neural Networks, Volume 4, IEEE, 1995, pp. 1942–1948.
[74] D. Cai, X. He, J. Han, Document clustering using locality preserving indexing, IEEE Trans. Knowl. Data Eng. 17 (2005) 1624–1637.
[75] L. Hubert, P. Arabie, Comparing partitions, J. Classification 2 (1985) 193–218.
[76] L. van der Maaten, G. Hinton, Visualizing data using t-SNE, J. Mach. Learn. Res. 9 (2008) 2579–2605.