Many-View Clustering - An Illustration using Multiple Dissimilarity Measures

Supplementary Material

Adán José-García
University of Manchester
adan.jose-garcia@manchester.ac.uk

Julia Handl
University of Manchester
julia.handl@manchester.ac.uk

Wilfrido Gómez-Flores
CINVESTAV-IPN
wgomez@tamps.cinvestav.mx

Mario Garza-Fabre
CINVESTAV-IPN
mgarza@tamps.cinvestav.mx

1 TABLES OF RESULTS

This appendix includes tables complementing the results of the experiments presented in the Section 4. The description of these synthetic and real-world datasets in terms of the number of patterns (N), number of dimensions (D) and number of clusters (K) are summarized in Table 1. The cluster structures of the synthetic datasets are illustrated in Figure 1. The Table 2 details the results of the ARI measure for the different clustering algorithms studied in this paper. The Table 3 presents the results in terms of the ARI measure for the kMOEA/D algorithm when the increasing the number of objectives (*i.e.*, views derived from the application of different dissimilarity measures).

Table 1: Description of the synthetic and real-world datasets.

Dataset	N	D	K	Clusters properties
Orange	400	2	2	Well-separated
Data_4_3	400	3	4	Well-separated
Lsun	400	2	3	Well-separated
Size5	1000	2	4	Overlapping
Data_5_2	250	2	5	Overlapping
Data_9_2	900	2	9	Overlapping
Inside	600	2	2	Nonlinearly separable
Atom	800	3	2	Nonlinearly separable
Spirals	1000	2	2	Nonlinearly separable
Iris	150	4	3	Overlapping classes
Wine	178	13	3	Overlapping classes
BreastCW	676	9	2	Overlapping classes
Thyroid	215	5	3	Overlapping classes and noise
Glass2	214	9	2	Overlapping classes and noise
Ecoli	336	7	8	Overlapping classes and noise

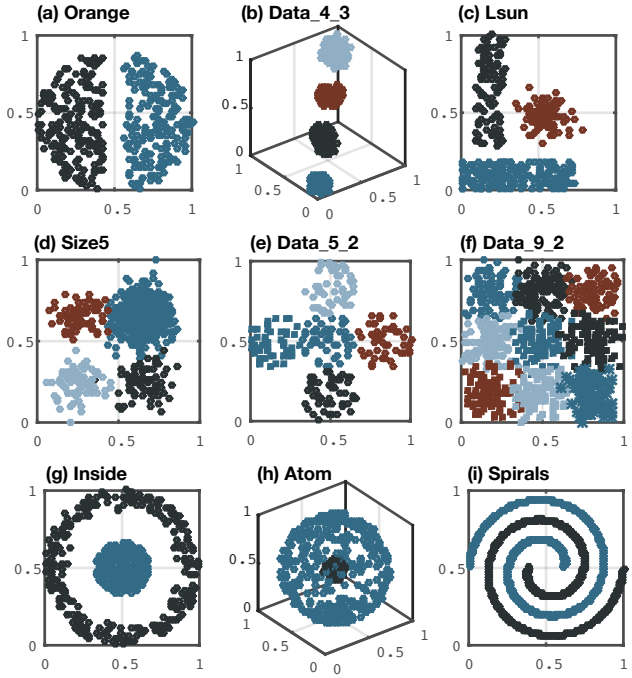

Figure 1: This figure illustrates the diversity of properties covered by our collection of test datasets: datasets (a) `Orange`, (b) `Data_4_2`, and (c) `Lsun` present well-separated clusters; datasets (d) `Size5`, (e) `Data_5_2`, and (f) `Data_9_2` exhibit overlapping clusters; and datasets (g) `Inside`, (h) `Atom`, and (i) `Spirals` consist of nonlinearly separable clusters.

GECCO '19 Companion, July 13–17, 2019, Prague, Czech Republic

ACM ISBN 978-1-4503-6748-6/19/07…$15.00
https://doi.org/10.1145/3319619.3323365

Table 2: ARI results obtained by different algorithms (mean values of 31 runs). The best average ARI value for each dataset has been shaded and bolded. The statistically best (α = 0.05) results are highlighted in bold face. The Kruskal–Wallis test with Bonferroni correction was applied to compare the algorithms. Dissimilarity measures: Euclidean distance (▲), MED distance based on Euclidean (△), Cosine distance (▼) MED distance based on cosine (▽).

Dataset	k-means			Single-Link			kDE			kMOEA/D		
	▲	△	▼	▲	△	▼	▲	△	▼	▲ △	▼▽	▲▼
Orange	**1.000 (0.00)**	**1.000 (0.00)**	0.389 (0.00)	**1.000**	**1.000**	0.000	**1.000 (0.00)**	**1.000 (0.00)**	0.389 (0.00)	**1.000 (0.00)**	0.428 (0.00)	**1.000 (0.00)**
Data_4_3	**1.000 (0.00)**	**1.000 (0.00)**	0.112 (0.00)	**1.000**	**1.000**	0.002	**1.000 (0.00)**	**1.000 (0.00)**	0.110 (0.01)	**1.000 (0.00)**	0.109 (0.04)	**1.000 (0.00)**
Lsun	0.436 (0.00)	**1.000 (0.00)**	0.535 (0.00)	**1.000**	**1.000**	0.391	0.463 (0.00)	**1.000 (0.00)**	0.546 (0.00)	**1.000 (0.00)**	0.552 (0.00)	0.914 (0.01)
Size5	0.920 (0.00)	0.739 (0.01)	0.332 (0.00)	0.002	0.015	0.032	**0.924 (0.00)**	0.711 (0.00)	0.308 (0.00)	**0.967 (0.00)**	0.779 (0.00)	**0.973 (0.00)**
Data_5_2	**0.888 (0.01)**	0.750 (0.02)	0.419 (0.01)	0.189	0.394	0.041	0.841 (0.01)	0.742 (0.01)	0.415 (0.00)	**0.943 (0.01)**	0.415 (0.01)	**0.971 (0.00)**
Data_9_2	**0.831 (0.00)**	0.506 (0.03)	0.293 (0.00)	0.000	0.000	0.010	**0.753 (0.04)**	0.512 (0.03)	0.294 (0.00)	**0.844 (0.00)**	0.315 (0.01)	**0.829 (0.01)**
Inside	0.001 (0.00)	**1.000 (0.00)**	0.016 (0.00)	**1.000**	**1.000**	0.000	0.001 (0.00)	**1.000 (0.00)**	0.002 (0.00)	**1.000 (0.00)**	0.144 (0.00)	0.003 (0.00)
Atom	0.184 (0.01)	**1.000 (0.00)**	0.136 (0.01)	**1.000**	**1.000**	0.000	0.097 (0.00)	**1.000 (0.00)**	0.149 (0.01)	**1.000 (0.00)**	0.196 (0.00)	0.197 (0.02)
Spirals	0.033 (0.00)	**1.000 (0.00)**	0.026 (0.00)	**1.000**	**1.000**	0.006	0.034 (0.00)	**1.000 (0.00)**	0.025 (0.00)	**1.000 (0.00)**	0.043 (0.00)	0.034 (0.00)
	0.59 (0.40)	**0.890 (0.17)**	0.250 (0.18)	0.690	0.710	0.050	0.570 (0.40)	**0.890 (0.17)**	0.250 (0.18)	**0.970 (0.05)**	0.330 (0.22)	0.660 (0.42)
0.66 (0.42)	0.730 (0.00)	0.717 (0.03)	**0.904 (0.00)**	0.561	0.564	0.566	0.746 (0.00)	0.755 (0.05)	0.886 (0.00)	**0.905 (0.00)**	**0.904 (0.00)**	**0.922 (0.00)**
Wine	**0.915 (0.00)**	0.550 (0.06)	**0.808 (0.00)**	0.004	0.006	0.004	0.756 (0.01)	0.711 (0.15)	0.706 (0.01)	0.786 (0.05)	0.735 (0.09)	0.770 (0.08)
BreastCW	0.861 (0.00)	0.747 (0.13)	0.856 (0.00)	0.007	0.003	0.003	0.856 (0.00)	0.804 (0.11)	**0.867 (0.00)**	0.833 (0.01)	**0.881 (0.00)**	**0.875 (0.00)**
Thyroid	**0.718 (0.00)**	**0.590 (0.14)**	0.186 (0.00)	0.430	0.232	0.013	**0.703 (0.00)**	0.220 (0.00)	0.137 (0.00)	**0.858 (0.02)**	0.199 (0.04)	**0.829 (0.05)**
Glass2	**0.634 (0.02)**	0.455 (0.05)	0.312 (0.00)	0.020	0.034	0.041	**0.587 (0.00)**	0.569 (0.00)	0.205 (0.03)	0.570 (0.01)	**0.780 (0.05)**	**0.642 (0.12)**
Ecoli	0.417 (0.03)	0.406 (0.04)	0.369 (0.02)	0.038	0.041	0.376	0.412 (0.03)	0.398 (0.05)	0.404 (0.02)	**0.531 (0.05)**	**0.699 (0.04)**	**0.445 (0.03)**
	0.710 (0.16)	0.580 (0.15)	0.570 (0.29)	0.170	0.140	0.160	0.680 (0.14)	0.580 (0.22)	0.530 (0.30)	**0.750 (0.15)**	**0.700 (0.24)**	**0.750 (0.17)**

Table 3: ARI results obtained by kMOEA/D when varying the number of objectives $\{2, 3, 4\}$. The best average ARI value for each dataset has been shaded and bolded. The statistically best (α = 0.05) results are highlighted in bold face. The Kruskal–Wallis test with Bonferroni correction was applied to compare the algorithms. Dissimilarity measures: Euclidean distance (▲), MED distance based on Euclidean (△), Cosine distance (▼) MED distance based on cosine (▽).

Dataset	2-objectives			3-objectives	4-objectives
	▲ △	▼▽	▲▼	▲ △ ▼	▲▼ △ ▽
Orange	**1.000 (0.00)**	0.428 (0.00)	**1.000 (0.00)**	**1.000 (0.00)**	**1.000 (0.00)**
Data_4_3	**1.000 (0.00)**	0.109 (0.04)	**1.000 (0.00)**	**1.000 (0.00)**	**1.000 (0.00)**
Lsun	**1.000 (0.00)**	0.552 (0.00)	0.914 (0.01)	**1.000 (0.00)**	**1.000 (0.00)**
Size5	0.967 (0.00)	0.779 (0.00)	**0.973 (0.00)**	**0.976 (0.00)**	**0.973 (0.00)**
Data_5_2	0.943 (0.01)	0.415 (0.01)	**0.971 (0.00)**	**0.970 (0.00)**	**0.961 (0.01)**
Data_9_2	**0.844 (0.00)**	0.315 (0.01)	0.829 (0.01)	**0.844 (0.01)**	**0.842 (0.01)**
Inside	**1.000 (0.00)**	0.144 (0.00)	0.003 (0.00)	**1.000 (0.00)**	**1.000 (0.00)**
Atom	**1.000 (0.00)**	0.196 (0.00)	0.197 (0.02)	**1.000 (0.00)**	**1.000 (0.00)**
Spirals	**1.000 (0.00)**	0.043 (0.00)	0.034 (0.00)	**1.000 (0.00)**	**1.000 (0.00)**
	0.973 (0.05)	0.331 (0.22)	0.658 (0.42)	**0.977 (0.05)**	**0.975 (0.05)**
Iris	0.905 (0.00)	0.904 (0.00)	0.922 (0.00)	**0.941 (0.00)**	0.928 (0.01)
Wine	0.786 (0.05)	0.735 (0.09)	0.770 (0.08)	**0.831 (0.03)**	**0.853 (0.01)**
BreastCW	0.833 (0.01)	0.881 (0.00)	0.875 (0.00)	0.879 (0.00)	**0.889 (0.00)**
Thyroid	0.858 (0.02)	0.199 (0.04)	0.829 (0.05)	**0.896 (0.01)**	**0.884 (0.01)**
Glass2	0.570 (0.01)	**0.780 (0.05)**	0.642 (0.12)	**0.748 (0.02)**	**0.761 (0.01)**
Ecoli	0.531 (0.05)	0.699 (0.04)	0.445 (0.03)	0.516 (0.01)	**0.753 (0.01)**
	0.747 (0.15)	0.700 (0.24)	0.743 (0.17)	**0.802 (0.14)**	**0.845 (0.07)**