

Supplementary Material

A General Framework for Mixed and Incomplete Data Clustering Based on Swarm Intelligence Algorithms

Yenny Villuendas-Rey ¹, Eley Barroso-Cubas ², Oscar Camacho-Nieto ^{1,*} and Cornelio Yáñez-Márquez ^{3,*}

¹ CIDETEC, Instituto Politécnico Nacional, Av. Juan de Dios Bátiz s/n, Nueva Industrial Vallejo, GAM, CDMX 07700, Mexico; yvilluendasr@ipn.mx

² Facultad de Ciencias Informáticas, Universidad de Ciego de Ávila, Modesto Reyes 65100, Cuba; eleybarroso@nauta.cu

³ CIC, Instituto Politécnico Nacional, Av. Juan de Dios Bátiz s/n, Nueva Industrial Vallejo, GAM, CDMX 07738, Mexico

* Correspondence: ocamacho@ipn.mx (O.C.-N.); coryanez@gmail.com (C.Y.-M.); Tel.: +52-5729600 (ext. 56584) (C.Y.-M.)

Citation: Villuendas-Rey, Y.; Barroso-Cubas, E.; Camacho-Nieto, O.; Yáñez-Márque, C. A General Framework for Mixed and Incomplete Data Clustering Based on Swarm Intelligence Algorithms. *Mathematics* **2021**, *9*, x. <http://doi.org/10.3390/math9070786>

Academic Editor: Yong-Hyuk Kim

Received: 16 February 2021

Accepted: 25 March 2021

Published: date

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

Experiment #1

Table S1. Cluster Error results of the CABC algorithm, with different iteration number.

Databases	50		100		500		1000	
	Avg	Stdev	Avg	Stdev	Avg	Stdev	Avg	Stdev
anneal	0.23	0.02	0.23	0.02	0.23	0.02	0.24	0.00
autos	0.56	0.03	0.56	0.03	0.56	0.03	0.55	0.03
cmc	0.56	0.02	0.56	0.02	0.56	0.02	0.57	0.00
colic	0.35	0.06	0.35	0.06	0.35	0.06	0.37	0.01
credit-a	0.40	0.08	0.40	0.08	0.40	0.08	0.42	0.03
credit-g	0.30	0.00	0.30	0.00	0.30	0.00	0.30	0.00
dermatology	0.46	0.08	0.46	0.08	0.46	0.08	0.40	0.08
heart-c	0.35	0.08	0.35	0.08	0.35	0.08	0.45	0.00
hepatitis	0.21	0.00	0.21	0.00	0.21	0.00	0.20	0.00
labor	0.33	0.03	0.33	0.03	0.33	0.03	0.31	0.05
lymph	0.36	0.05	0.36	0.05	0.36	0.05	0.40	0.05
postoperative	0.36	0.00	0.36	0.00	0.36	0.00	0.35	0.00
tae	0.60	0.03	0.60	0.03	0.60	0.03	0.61	0.03
vowel	0.83	0.03	0.83	0.03	0.83	0.03	0.83	0.02
zoo	0.21	0.05	0.21	0.05	0.21	0.05	0.28	0.08

Table S2. Cluster Error results of the CFA algorithm, with different iteration number.

Databases	50		100		500		1000	
	Avg	Stdev	Avg	Stdev	Avg	Stdev	Avg	Stdev
anneal	0.23	0.02	0.23	0.02	0.23	0.02	0.24	0.00
autos	0.55	0.03	0.55	0.03	0.55	0.03	0.57	0.04
cmc	0.57	0.00	0.57	0.00	0.57	0.00	0.57	0.01
colic	0.32	0.06	0.32	0.06	0.32	0.06	0.37	0.00
credit-a	0.44	0.01	0.44	0.01	0.44	0.01	0.43	0.02
credit-g	0.30	0.00	0.30	0.00	0.30	0.00	0.30	0.00
dermatology	0.48	0.06	0.48	0.06	0.48	0.06	0.49	0.10
heart-c	0.38	0.06	0.38	0.06	0.38	0.06	0.40	0.07
hepatitis	0.21	0.00	0.21	0.00	0.21	0.00	0.20	0.02
labor	0.30	0.05	0.30	0.05	0.30	0.05	0.33	0.04
lymph	0.37	0.04	0.37	0.04	0.37	0.04	0.38	0.06
postoperative	0.36	0.01	0.36	0.01	0.36	0.01	0.34	0.03
tae	0.54	0.05	0.54	0.05	0.54	0.05	0.61	0.02
vowel	0.82	0.02	0.82	0.02	0.82	0.02	0.80	0.02
zoo	0.24	0.07	0.24	0.07	0.24	0.07	0.24	0.08

Table S3. Cluster Error results of the CNBA algorithm, with different iteration number.

Databases	50		100		500		1000	
	Avg	Stdev	Avg	Stdev	Avg	Stdev	Avg	Stdev
anneal	0.24	0.00	0.24	0.00	0.24	0.00	0.24	0.00
autos	0.56	0.03	0.56	0.03	0.56	0.03	0.56	0.04
cmc	0.57	0.01	0.57	0.01	0.57	0.01	0.57	0.01
colic	0.35	0.05	0.35	0.05	0.35	0.05	0.34	0.06
credit-a	0.41	0.06	0.41	0.06	0.41	0.06	0.36	0.09
credit-g	0.30	0.00	0.30	0.00	0.30	0.00	0.30	0.00
dermatology	0.46	0.05	0.46	0.05	0.46	0.05	0.47	0.07
heart-c	0.33	0.07	0.33	0.07	0.33	0.07	0.35	0.07
hepatitis	0.21	0.00	0.21	0.00	0.21	0.00	0.21	0.00
labor	0.33	0.03	0.33	0.03	0.33	0.03	0.34	0.02
lymph	0.37	0.06	0.37	0.06	0.37	0.06	0.37	0.06
postoperative	0.36	0.00	0.36	0.00	0.36	0.00	0.35	0.01

tae	0.61	0.02	0.61	0.02	0.61	0.02	0.58	0.02
vowel	0.83	0.02	0.83	0.02	0.83	0.02	0.82	0.01
zoo	0.27	0.07	0.27	0.07	0.27	0.07	0.27	0.07

Table S3. Results of Wilcoxon test comparing the results of the proposed algorithms with different iteration number, according to Cluster Error.

Algorithm	Pair	Cluster Error		
		win-loses-ties	p-value	Decision
CABC	50 it vs. 100 it	5-9-1	0.198	Do not reject H0
	50 it vs. 500 it	6-9-0	0.865	Do not reject H0
	50 it vs. 1000 it	6-9-0	0.609	Do not reject H0
CFA	50 it vs. 100 it	5-7-3	0.754	Do not reject H0
	50 it vs. 500 it	5-7-3	0.754	Do not reject H0
	50 it vs. 1000 it	4-10-1	0.551	Do not reject H0
CNBA	50 it vs. 100 it	7-7-1	0.594	Do not reject H0
	50 it vs. 500 it	6-7-2	0.917	Do not reject H0
	50 it vs. 1000 it	7-8-0	0.955	Do not reject H0

The Friedman test comparing the different iteration number for CABC, CFA and CNBA algorithms according to Cluster Error obtained p-values of 0.923263, 0.391625 and 0.989334, respectively. Due to no differences in the performance of the algorithms were found, there is no need to apply post-hoc tests.

Table S4. Adjusted Rand Index results of the CABC algorithm, with different iteration number.

Datasets	50		100		500		1000	
	Avg	Stdev	Avg	Stdev	Avg	Stdev	Avg	Stdev
anneal	0.06	0.08	0.08	0.08	0.07	0.07	0.05	0.04
autos	0.08	0.03	0.11	0.03	0.12	0.04	0.09	0.03
cmc	0.01	0.01	0.01	0.01	0.01	0.02	0.00	0.01
colic	0.06	0.12	0.09	0.15	0.08	0.13	0.01	0.03
credit-a	0.06	0.11	0.07	0.16	0.00	0.01	0.02	0.02
credit-g	0.02	0.03	0.02	0.05	0.00	0.01	0.00	0.02
dermatology	0.27	0.12	0.36	0.10	0.33	0.11	0.35	0.14
heart-c	0.10	0.09	0.04	0.06	0.01	0.03	0.00	0.00
hepatitis	0.01	0.08	-0.01	0.04	-0.02	0.04	-0.01	0.03
labor	0.04	0.08	0.20	0.23	0.11	0.18	0.08	0.12
lymph	0.09	0.07	0.09	0.04	0.07	0.05	0.08	0.07
postoperative	0.01	0.02	0.03	0.04	0.01	0.04	0.03	0.02
tae	0.01	0.02	0.02	0.02	0.01	0.01	0.01	0.01
vowel	0.02	0.01	0.01	0.01	0.03	0.03	0.03	0.02
zoo	0.61	0.15	0.48	0.21	0.48	0.17	0.47	0.14

Table S5. Adjusted Rand Index results of the CFA algorithm, with different iteration number.

Datasets	50		100		500		1000	
	Avg	Stdev	Avg	Stdev	Avg	Stdev	Avg	Stdev
anneal	0.02	0.05	0.02	0.03	0.01	0.03	0.00	0.03
autos	0.08	0.03	0.09	0.03	0.11	0.03	0.08	0.04
cmc	0.00	0.00	0.01	0.01	0.00	0.01	0.01	0.01
colic	0.11	0.14	0.10	0.09	0.01	0.08	-0.03	0.00
credit-a	0.01	0.01	0.01	0.02	0.06	0.15	0.01	0.02
credit-g	0.01	0.03	0.01	0.02	0.01	0.02	0.01	0.02
dermatology	0.19	0.08	0.25	0.15	0.22	0.13	0.19	0.11
heart-c	0.06	0.06	0.05	0.08	0.07	0.06	0.06	0.09
hepatitis	0.01	0.00	0.01	0.01	0.01	0.02	0.04	0.13
labor	0.13	0.10	0.15	0.10	0.11	0.11	0.05	0.09

lymph	0.06	0.04	0.04	0.03	0.08	0.06	0.06	0.05
postoperative	0.02	0.02	0.01	0.03	0.03	0.03	0.03	0.03
tae	0.04	0.04	0.01	0.02	0.02	0.02	0.00	0.01
vowel	0.03	0.02	0.03	0.02	0.03	0.01	0.04	0.01
zoo	0.52	0.18	0.51	0.17	0.53	0.12	0.48	0.16

Table S6. Adjusted Rand Index results of the CNBA algorithm, with different iteration number.

Datasets	50		100		500		1000	
	Avg	Stdev	Avg	Stdev	Avg	Stdev	Avg	Stdev
anneal	0.05	0.05	0.02	0.02	0.00	0.02	0.00	0.03
autos	0.10	0.04	0.07	0.02	0.09	0.03	0.08	0.02
cmc	0.00	0.01	0.00	0.01	0.01	0.01	0.01	0.02
colic	0.07	0.13	0.05	0.10	0.07	0.09	0.06	0.09
credit-a	0.11	0.15	0.04	0.07	0.01	0.01	0.01	0.01
credit-g	0.01	0.03	0.02	0.02	0.02	0.02	0.00	0.02
dermatology	0.23	0.09	0.21	0.08	0.26	0.16	0.24	0.07
heart-c	0.10	0.08	0.13	0.08	0.02	0.03	0.09	0.10
hepatitis	0.02	0.10	-0.01	0.05	0.02	0.06	0.00	0.04
labor	0.03	0.05	0.03	0.08	0.10	0.09	0.08	0.08
lymph	0.07	0.04	0.06	0.04	0.07	0.06	0.07	0.06
postoperative	0.02	0.03	0.01	0.02	0.01	0.02	0.02	0.03
tae	0.02	0.01	0.00	0.01	0.01	0.02	0.01	0.01
vowel	0.03	0.01	0.02	0.01	0.04	0.02	0.02	0.02
zoo	0.52	0.21	0.46	0.15	0.47	0.10	0.54	0.22

Experiment #2

Table S7. Cluster Error results of the CABC algorithm, with different population sizes.

Databases	10		20		30		40	
	Avg	Stdev	Avg	Stdev	Avg	Stdev	Avg	Stdev
anneal	0.23	0.02	0.23	0.02	0.23	0.01	0.24	0.00
autos	0.56	0.03	0.56	0.04	0.53	0.04	0.54	0.02
cmc	0.56	0.02	0.56	0.01	0.56	0.02	0.57	0.01
colic	0.35	0.06	0.37	0.00	0.36	0.03	0.33	0.06
credit-a	0.40	0.08	0.42	0.02	0.43	0.04	0.40	0.04
credit-g	0.30	0.00	0.30	0.00	0.30	0.00	0.30	0.00
dermatology	0.46	0.08	0.44	0.09	0.43	0.06	0.41	0.11
heart-c	0.35	0.08	0.40	0.05	0.43	0.05	0.41	0.08
hepatitis	0.21	0.00	0.21	0.00	0.20	0.00	0.20	0.00
labor	0.33	0.03	0.34	0.03	0.33	0.04	0.31	0.04
lymph	0.36	0.05	0.40	0.06	0.37	0.07	0.38	0.05
postoperative	0.36	0.00	0.36	0.00	0.35	0.00	0.36	0.00
tae	0.60	0.03	0.59	0.04	0.58	0.02	0.61	0.04
vowel	0.83	0.03	0.84	0.02	0.82	0.03	0.84	0.03
zoo	0.21	0.05	0.25	0.07	0.26	0.06	0.28	0.05

Table S8. Cluster Error results of the CFA algorithm, with different population sizes.

Databases	10		20		30		40	
	Avg	Stdev	Avg	Stdev	Avg	Stdev	Avg	Stdev
anneal	0.23	0.02	0.24	0.00	0.24	0.00	0.24	0.00
autos	0.55	0.03	0.54	0.05	0.56	0.04	0.53	0.05
cmc	0.57	0.00	0.57	0.00	0.57	0.01	0.56	0.01
colic	0.32	0.06	0.33	0.04	0.36	0.03	0.35	0.06
credit-a	0.44	0.01	0.41	0.09	0.43	0.03	0.44	0.01
credit-g	0.30	0.00	0.30	0.00	0.30	0.00	0.30	0.00

dermatology	0.48	0.06	0.50	0.10	0.42	0.09	0.47	0.06
heart-c	0.38	0.06	0.37	0.08	0.37	0.08	0.37	0.07
hepatitis	0.21	0.00	0.21	0.00	0.21	0.00	0.21	0.00
labor	0.30	0.05	0.32	0.05	0.31	0.04	0.29	0.07
lymph	0.37	0.04	0.37	0.04	0.39	0.05	0.36	0.06
postoperative	0.36	0.01	0.35	0.01	0.35	0.01	0.35	0.00
tae	0.54	0.05	0.57	0.03	0.57	0.08	0.55	0.10
vowel	0.82	0.02	0.82	0.03	0.82	0.01	0.83	0.03
zoo	0.24	0.07	0.29	0.07	0.28	0.07	0.27	0.08

Table S9. Cluster Error results of the CNBA algorithm, with different population sizes.

Databases	10		20		30		40	
	Avg	Stdev	Avg	Stdev	Avg	Stdev	Avg	Stdev
anneal	0.24	0.00	0.24	0.00	0.24	0.00	0.24	0.00
autos	0.56	0.03	0.55	0.03	0.55	0.03	0.56	0.04
cmc	0.57	0.01	0.56	0.01	0.57	0.00	0.57	0.01
colic	0.35	0.05	0.37	0.00	0.34	0.06	0.37	0.01
credit-a	0.41	0.06	0.40	0.09	0.41	0.06	0.42	0.02
credit-g	0.30	0.00	0.30	0.00	0.30	0.00	0.30	0.00
dermatology	0.46	0.05	0.52	0.07	0.46	0.07	0.42	0.07
heart-c	0.33	0.07	0.37	0.07	0.39	0.08	0.40	0.05
hepatitis	0.21	0.00	0.20	0.00	0.21	0.00	0.21	0.00
labor	0.33	0.03	0.33	0.04	0.34	0.02	0.32	0.04
lymph	0.37	0.06	0.39	0.04	0.37	0.07	0.37	0.05
postoperative	0.36	0.00	0.35	0.00	0.36	0.00	0.35	0.01
tae	0.61	0.02	0.57	0.03	0.60	0.02	0.57	0.04
vowel	0.83	0.02	0.83	0.02	0.81	0.02	0.79	0.03
zoo	0.27	0.07	0.27	0.10	0.27	0.08	0.31	0.09

Table S10. Results of Wilcoxon test comparing the results of the proposed algorithms with different population size.

Algorithm	Pair	Cluster Error		
		win-loses-ties	p-value	Decision
CABC	10 pop vs. 20 pop	7-5-3	0.388	Do not reject H0
	10 pop vs. 30 pop	9-5-1	0.198	Do not reject H0
	10 pop vs. 40 pop	7-6-2	0.507	Do not reject H0
CFA	10 pop vs. 20 pop	7-6-2	0.753	Do not reject H0
	10 pop vs. 30 pop	6-6-3	0.695	Do not reject H0
	10 pop vs. 40 pop	4-11-0	0.173	Do not reject H0
CNBA	10 pop vs. 20 pop	6-7-2	0.701	Do not reject H0
	10 pop vs. 30 pop	7-6-2	0.861	Do not reject H0
	10 pop vs. 40 pop	7-7-1	0.778	Do not reject H0

The Friedman test obtained p-values of 0.944342, 0.628078, and 0.989334, respectively. Due to no differences in the performance of the algorithms were found, there is no need to apply post-hoc tests.

Table S11. Adjusted Rand Index results of the CABC algorithm, with different population sizes.

Databases	10		20		30		40	
	Avg	Stdev	Avg	Stdev	Avg	Stdev	Avg	Stdev
anneal	0.06	0.08	0.08	0.09	0.06	0.05	0.02	0.04
autos	0.08	0.03	0.08	0.05	0.10	0.03	0.11	0.03
cmc	0.01	0.01	0.00	0.01	0.01	0.02	0.01	0.02
colic	0.06	0.12	0.00	0.03	0.02	0.06	0.08	0.12
credit-a	0.06	0.11	0.02	0.02	0.02	0.03	0.04	0.04
credit-g	0.02	0.03	0.02	0.02	0.01	0.01	0.01	0.03

dermatology	0.27	0.12	0.27	0.13	0.28	0.10	0.32	0.15
heart-c	0.10	0.09	0.04	0.05	0.02	0.04	0.05	0.10
hepatitis	0.01	0.08	0.03	0.06	0.00	0.05	0.02	0.04
labor	0.04	0.08	0.04	0.07	0.07	0.08	0.10	0.10
lymph	0.09	0.07	0.06	0.06	0.09	0.05	0.08	0.05
postoperative	0.01	0.02	0.03	0.03	0.00	0.02	0.01	0.02
tae	0.01	0.02	0.02	0.02	0.02	0.01	0.01	0.02
vowel	0.02	0.01	0.02	0.01	0.03	0.01	0.02	0.02
zoo	0.61	0.15	0.53	0.20	0.51	0.16	0.51	0.13

Table S12. Adjusted Rand Index results of the CFA algorithm, with different population sizes.

Databases	10		20		30		40	
	Avg	Stdev	Avg	Stdev	Avg	Stdev	Avg	Stdev
anneal	0.02	0.05	0.02	0.05	0.04	0.04	0.00	0.02
autos	0.08	0.03	0.08	0.04	0.09	0.03	0.11	0.06
cmc	0.00	0.00	0.00	0.01	0.01	0.01	0.02	0.01
colic	0.11	0.14	0.07	0.11	0.00	0.06	0.01	0.13
credit-a	0.01	0.01	0.06	0.16	0.01	0.03	0.01	0.01
credit-g	0.01	0.03	0.00	0.01	0.01	0.02	0.02	0.03
dermatology	0.19	0.08	0.20	0.15	0.28	0.15	0.22	0.11
heart-c	0.06	0.06	0.08	0.09	0.09	0.09	0.08	0.07
hepatitis	0.01	0.00	-0.01	0.04	-0.01	0.05	-0.01	0.03
labor	0.13	0.10	0.08	0.11	0.09	0.10	0.14	0.16
lymph	0.06	0.04	0.08	0.05	0.07	0.04	0.10	0.06
postoperative	0.02	0.02	0.03	0.03	0.01	0.02	0.02	0.04
tae	0.04	0.04	0.02	0.02	0.05	0.11	0.06	0.10
vowel	0.03	0.02	0.02	0.02	0.02	0.01	0.02	0.01
zoo	0.52	0.18	0.46	0.15	0.52	0.20	0.52	0.19

Table S13. Adjusted Rand Index results of the CNBA algorithm, with different population sizes.

Databases	10		20		30		40	
	Avg	Stdev	Avg	Stdev	Avg	Stdev	Avg	Stdev
anneal	0.05	0.05	0.02	0.04	0.04	0.08	0.07	0.06
autos	0.10	0.04	0.08	0.03	0.07	0.03	0.08	0.03
cmc	0.00	0.01	0.01	0.01	0.01	0.01	0.01	0.01
colic	0.07	0.13	0.02	0.03	0.07	0.13	0.00	0.03
credit-a	0.11	0.15	0.06	0.16	0.04	0.07	0.02	0.02
credit-g	0.01	0.03	0.01	0.02	0.01	0.01	0.01	0.03
dermatology	0.23	0.09	0.17	0.08	0.23	0.10	0.29	0.14
heart-c	0.10	0.08	0.08	0.09	0.07	0.09	0.04	0.05
hepatitis	0.02	0.10	0.00	0.06	0.01	0.03	-0.02	0.04
labor	0.03	0.05	0.08	0.10	0.04	0.05	0.09	0.10
lymph	0.07	0.04	0.05	0.04	0.10	0.11	0.08	0.06
postoperative	0.02	0.03	0.02	0.04	0.01	0.02	0.01	0.03
tae	0.02	0.01	0.02	0.02	0.01	0.01	0.02	0.02
vowel	0.03	0.01	0.02	0.01	0.03	0.01	0.03	0.02
zoo	0.52	0.21	0.49	0.24	0.47	0.20	0.36	0.20

Experiment #3

Table S14. Cluster Error results of the CABC algorithm, with different fitness functions.

Databases	Davies - Bouldin		Dunn's		Silhouette	
	Avg	Stdev	Avg	Stdev	Avg	Stdev
anneal	0.23	0.02	0.23	0.01	0.21	0.02
autos	0.56	0.03	0.53	0.04	0.56	0.03
cmc	0.56	0.02	0.57	0.00	0.57	0.00
colic	0.35	0.06	0.36	0.01	0.35	0.03
credit-a	0.40	0.08	0.40	0.03	0.44	0.00
credit-g	0.30	0.00	0.30	0.00	0.30	0.00
dermatology	0.46	0.08	0.28	0.07	0.28	0.06
heart-c	0.35	0.08	0.19	0.02	0.18	0.01
hepatitis	0.21	0.00	0.20	0.01	0.21	0.00
labor	0.33	0.03	0.32	0.03	0.32	0.04
lymph	0.36	0.05	0.30	0.05	0.35	0.07
postoperative	0.36	0.00	0.35	0.01	0.35	0.00
tae	0.60	0.03	0.57	0.01	0.56	0.03
vowel	0.83	0.03	0.86	0.01	0.89	0.01
zoo	0.21	0.05	0.15	0.05	0.17	0.05

Table S15. Cluster Error results of the CFA algorithm, with different fitness functions.

Databases	Davies - Bouldin		Dunn's		Silhouette	
	Avg	Stdev	Avg	Stdev	Avg	Stdev
anneal	0.23	0.02	0.23	0.01	0.21	0.03
autos	0.55	0.03	0.54	0.02	0.54	0.07
cmc	0.57	0.00	0.57	0.00	0.57	0.00
colic	0.32	0.06	0.37	0.00	0.35	0.03
credit-a	0.44	0.01	0.42	0.04	0.44	0.00
credit-g	0.30	0.00	0.30	0.00	0.30	0.00
dermatology	0.48	0.06	0.33	0.04	0.29	0.05
heart-c	0.38	0.06	0.21	0.02	0.19	0.01
hepatitis	0.21	0.00	0.20	0.01	0.20	0.00
labor	0.30	0.05	0.35	0.01	0.32	0.04
lymph	0.37	0.04	0.32	0.06	0.34	0.06
postoperative	0.36	0.01	0.36	0.00	0.35	0.00
tae	0.54	0.05	0.57	0.00	0.56	0.03
vowel	0.82	0.02	0.86	0.01	0.90	0.01
zoo	0.24	0.07	0.09	0.03	0.15	0.06

Table S16. Cluster Error results of the CNBA algorithm, with different fitness functions.

Databases	Davies - Bouldin		Dunn's		Silhouette	
	Avg	Stdev	Avg	Stdev	Avg	Stdev
anneal	0.24	0.00	0.22	0.02	0.22	0.02
autos	0.56	0.03	0.54	0.04	0.55	0.04
cmc	0.57	0.01	0.57	0.00	0.57	0.00
colic	0.35	0.05	0.37	0.00	0.35	0.03
credit-a	0.41	0.06	0.40	0.07	0.44	0.00
credit-g	0.30	0.00	0.30	0.00	0.30	0.00
dermatology	0.46	0.05	0.31	0.05	0.26	0.11
heart-c	0.33	0.07	0.20	0.02	0.18	0.01
hepatitis	0.21	0.00	0.20	0.01	0.20	0.01
labor	0.33	0.03	0.34	0.02	0.30	0.06
lymph	0.37	0.06	0.33	0.04	0.28	0.05
postoperative	0.36	0.00	0.35	0.01	0.35	0.00

tae	0.61	0.02	0.56	0.02	0.56	0.02
vowel	0.83	0.02	0.85	0.01	0.89	0.01
zoo	0.27	0.07	0.11	0.02	0.18	0.06

Table S17. Results of Wilcoxon test comparing the results of the proposed algorithms with different fitness function.

Algorithm	Pair	Cluster Error		
		win-loses-ties	p-value	Decision
CABC	Davies - Bouldin vs. Dunn's	5-9-1	0.394	Do not reject H0
	Davies - Bouldin vs. Silhouette	6-8-1	0.433	Do not reject H0
	Dunn's vs. Silhouette	8-7-0	0.925	Do not reject H0
CFA	Davies - Bouldin vs. Dunn's	8-7-0	0.363	Do not reject H0
	Davies - Bouldin vs. Silhouette	7-8-0	0.532	Do not reject H0
	Dunn's vs. Silhouette	5-10-0	0.776	Do not reject H0
CNBA	Davies - Bouldin vs. Dunn's	5-8-2	0.730	Do not reject H0
	Davies - Bouldin vs. Silhouette	5-8-2	0.507	Do not reject H0
	Dunn's vs. Silhouette	6-8-1	0.382	Do not reject H0

The Friedman test obtained p-values of 0.361799, 0.627089, and 0.165299, respectively. Due to no differences in the performance of the algorithms were found, there is no need to apply post-hoc tests.

Table S18. Adjusted Rand Index results of the CABC algorithm, with different fitness functions.

Databases	Davies - Bouldin		Dunn's		Silhouette	
	Avg	Stdev	Avg	Stdev	Avg	Stdev
anneal	0.06	0.08	0.09	0.06	0.12	0.08
autos	0.08	0.03	0.11	0.03	0.10	0.03
cmc	0.01	0.01	0.00	0.01	0.01	0.01
colic	0.06	0.12	0.05	0.03	0.07	0.07
credit-a	0.06	0.11	0.04	0.03	0.00	0.00
credit-g	0.02	0.03	0.03	0.03	0.03	0.03
dermatology	0.27	0.12	0.47	0.12	0.48	0.09
heart-c	0.10	0.09	0.38	0.04	0.40	0.04
hepatitis	0.01	0.08	0.15	0.06	0.15	0.13
labor	0.04	0.08	0.08	0.08	0.05	0.09
lymph	0.09	0.07	0.13	0.06	0.12	0.10
postoperative	0.01	0.02	0.03	0.04	0.02	0.04
tae	0.01	0.02	0.03	0.01	0.03	0.02
vowel	0.02	0.01	0.01	0.01	0.00	0.00
zoo	0.61	0.15	0.75	0.11	0.72	0.13

Table S19. Adjusted Rand Index results of the CFA algorithm, with different fitness functions.

Databases	Davies - Bouldin		Dunn's		Silhouette	
	Avg	Stdev	Avg	Stdev	Avg	Stdev
anneal	0.02	0.05	0.06	0.04	0.12	0.10
autos	0.08	0.03	0.09	0.02	0.10	0.07
cmc	0.00	0.00	-0.01	0.01	0.01	0.01
colic	0.11	0.14	0.05	0.01	0.05	0.08
credit-a	0.01	0.01	0.02	0.03	0.01	0.00
credit-g	0.01	0.03	0.04	0.02	0.03	0.01
dermatology	0.19	0.08	0.41	0.08	0.49	0.09
heart-c	0.06	0.06	0.35	0.05	0.39	0.02
hepatitis	0.01	0.00	0.19	0.01	0.01	0.04
labor	0.13	0.10	0.00	0.02	0.04	0.08
lymph	0.06	0.04	0.13	0.05	0.15	0.07

postoperative	0.02	0.02	0.02	0.03	0.00	0.04
tae	0.04	0.04	0.02	0.00	0.03	0.02
vowel	0.03	0.02	0.01	0.01	-0.01	0.00
zoo	0.52	0.18	0.84	0.07	0.81	0.10

Table S20. Adjusted Rand Index results of the CNBA algorithm, with different fitness functions.

Databases	Davies - Bouldin		Dunn's		Silhouette	
	Avg	Stdev	Avg	Stdev	Avg	Stdev
anneal	0.05	0.05	0.07	0.07	0.07	0.06
autos	0.10	0.04	0.10	0.03	0.11	0.03
cmc	0.00	0.01	-0.01	0.01	0.01	0.01
colic	0.07	0.13	0.04	0.03	0.06	0.08
credit-a	0.11	0.15	0.05	0.08	0.00	0.00
credit-g	0.01	0.03	0.02	0.02	0.03	0.02
dermatology	0.23	0.09	0.45	0.09	0.54	0.17
heart-c	0.10	0.08	0.37	0.05	0.41	0.02
hepatitis	0.02	0.10	0.17	0.06	0.05	0.12
labor	0.03	0.05	0.04	0.06	0.10	0.13
lymph	0.07	0.04	0.11	0.05	0.17	0.06
postoperative	0.02	0.03	0.06	0.02	0.00	0.02
tae	0.02	0.01	0.03	0.01	0.03	0.02
vowel	0.03	0.01	0.01	0.01	0.00	0.00
zoo	0.52	0.21	0.74	0.10	0.69	0.18

Experiment #4

Table S20. Results of Wilcoxon test comparing the results of the proposed algorithms with respect to other clustering algorithms, according to Cluster Error.

Algorithm	versus	Cluster Error		
		win-loses-ties	p-value	Decision
CABC	kPrototypes	12-3-0	0.023	Reject H0
	AGKA	15-0-0	0.001	Reject H0
	AD2011	13-2-0	0.003	Reject H0
CFA	kPrototypes	12-3-0	0.031	Reject H0
	AGKA	12-3-0	0.005	Reject H0
	AD2011	13-2-0	0.003	Reject H0
CNBA	kPrototypes	11-3-1	0.019	Reject H0
	AGKA	14-1-0	0.001	Reject H0
	AD2011	12-2-1	0.002	Reject H0