WEB SUPPLEMENT

SUPPLEMENT 1 ONLINE

Description of the look-ahead algorithm. All variables are global except for the local $\mathrm{A}, \mathrm{A}_1, \mathrm{A}_2, \mathrm{A}_1^*, \mathrm{A}_2^*, \mathrm{parent}, \mathrm{g}$ and $\mathrm{pointer}_2$. The initial setup is: $\mathrm{generation\_index} = 0$, $\mathrm{first\_split} = \mathrm{TRUE}$, $\mathrm{A} = \mathrm{root}$. The user sets the initial positions of the flat clusters, Y and the number of steps, h, to look ahead.

procedure *look_ahead* $(\mathrm{A}, \mathrm{h}, \mathrm{Y})$; Y global;
if A is a leaf
then
if $\mathrm{A} = \mathrm{root}$
then
link(A, Y)
else
if $\mathrm{case} \in \{\mathrm{C}, \mathrm{D}\}$
then
link(A, Y)
else
find the two descendants of $\mathrm{A}: \mathrm{A}_1$ and A_2
$\mathrm{generation\_index} = \mathrm{generation\_index} + 1$
compute $\sigma = \mathrm{S}(\mathrm{A}), \sigma_1 = \mathrm{S}'(\mathrm{A}_1, \mathrm{A}_2)$ and $\sigma_2 = \mathrm{S}'(\mathrm{A}_2, \mathrm{A}_1)$
$\mathrm{Y}_{12} = \mathrm{gravity\_center}((\mathrm{A}_1, \mathrm{A}_2), \mathrm{Y})$
$\mathrm{Y}_{21} = \mathrm{gravity\_center}((\mathrm{A}_2, \mathrm{A}_1), \mathrm{Y})$
if $\max(\sigma_1, \sigma_2) < \sigma$
then
if $\mathrm{generation\_index} > \mathrm{h}$
then
$\mathrm{case} = A$
case_A(A)
else
$\mathrm{case} = B$
case_B$(\mathrm{A}, \mathrm{A}_1, \mathrm{A}_2)$
else
if $\mathrm{first\_split} = \mathrm{TRUE}$
then
$\mathrm{case} = C$
case_C$(\mathrm{A}, \mathrm{A}_1, \mathrm{A}_2)$
else
$\mathrm{case} = D$
case_D()

procedure *case_A*(A)
% we cannot split in the h-th step, then we skip back to the
% (h-1)-th step
link(A, Y)
generation_index = generation_index - 1

procedure *case_B*$(\mathrm{A}, \mathrm{A}_1, \mathrm{A}_2)$
% we cannot split but we still can look some step ahead; the
% first time the procedure is called, a global pointer is set at
% the explored branch
if$(\sigma_1 > \sigma_2)$
then
A_1^*= A_1
A_2^*= A_2
Y= Y_{12}
else
A_1^*= A_2
A_2^*= A_1
Y = Y_{21}
if first_split = TRUE
then
first_split = FALSE
pointer = A
$\mathrm{Y_p} = \mathrm{Y}$
parent = A
descendants = $(\mathrm{A}_1^*, \mathrm{A}_2^*)$
if $(\sigma_1 > \sigma_2)$
then
$\mathrm{Y_d} = \mathrm{Y}_{12}$
else
$\mathrm{Y_d} = \mathrm{Y}_{21}$
if A_1^* and A_2^* are leaves
then
link$(\mathrm{A}, \mathrm{Y_p})$
else
g = generation_index
link($(\mathrm{A}_1^*, \mathrm{A}_2^*)$, Y)
look_ahead(A_1^*, Y)
if pointer = parent
then
generation_index = g
look_ahead(A_2^*, Y)
if pointer = parent
then
link(A, Y)

procedure *case_C*$(\mathrm{A}, \mathrm{A}_1, \mathrm{A}_2)$
% split normally
generation_index = 0
if $(\sigma_1 > \sigma_2)$
then
$\mathrm{A}_1^* = \mathrm{A}_1$
$\mathrm{A}_2^* = \mathrm{A}_2$
$\mathrm{Y} = \mathrm{Y}_{12}$
else
$\mathrm{A}_1^* = \mathrm{A}_2$
$\mathrm{A}_2^* = \mathrm{A}_1$
$\mathrm{Y} = \mathrm{Y}_{21}$
$\mathrm{pointer} = \mathrm{A}_1^*$
link$((\mathrm{A}_1^*, \mathrm{A}_2^*), \mathrm{Y})$
look_ahead$(\mathrm{A}_1^*, \mathrm{Y})$
$\mathrm{pointer} = \mathrm{A}_2^*$
generation_index = 0
first_split = TRUE
look_ahead$(\mathrm{A}_2^*, \mathrm{Y})$

procedure *case_D*()
% split the branch marked with the global pointer
$\mathrm{Y} = \mathrm{Y_d}$
link(descendants, Y)
pointer = descendants(1)
$\mathrm{pointer}_2$ = descendants(2)
generation_index = 0

first_split = TRUE
look_ahead(pointer, Y)
generation_index = 0
$\mathrm{pointer} = \mathrm{pointer}_2$
first_split = TRUE
look_ahead(pointer, Y)

SUPPLEMENT 2 ONLINE

Table 2. Comparison between a hierarchical and a non-hierarchical clustering, with no look ahead, computing the information theoretical-based score S_1.

	Number of original groups	Noise B/W	% elements closer to other centres than their own	Clustering 1 % of misclassified elements		Clustering 2 % of misclassified elements		Average number of superclusters	Average number of branches in the tree
				Hierarchical tree	Grouping	k-means (k=10)	Grouping		
	4	9.17	0.00%	3.93%	3.93%	0.76%	2.55%	3.80	3.91
	4	4.36	0.00%	5.72%	5.72%	0.69%	4.45%	3.66	3.79
Cluster of	4	3.25	0.29%	9.45%	9.45%	0.57%	4.03%	3.50	3.66
different size	4	2.15	2.61%	20.38%	21.01%	2.38%	11.28%	2.89	3.22
	4	1.84	5.26%	31.86%	31.92%	6.04%	17.68%	2.23	2.44
	4	1.38	6.11%	50.11%	50.38%	10.72%	19.74%	1.46	1.62
	8	8.95	0.00%	9.28%	9.86%	5.53%	9.19%	6.57	5.27
	8	4.67	0.06%	12.48%	13.35%	5.91%	9.50%	6.33	5.32
Cluster of	8	2.59	0.89%	14.93%	15.74%	3.94%	10.14%	6.08	5.20
different size	8	2.18	6.34%	33.71%	34.48%	6.13%	24.41%	4.44	4.73
	8	1.58	7.17%	59.78%	60.34%	10.61%	25.41%	2.27	2.69
	8	1.09	13.13%	75.49%	75.50%	20.46%	28.78%	1.20	1.27

Table 3. Comparison between a hierarchical and a non-hierarchical clustering, using the look ahead algorithm, searching one step ahead and computing the information theoretical-based score S_1.

	Number of original groups	Noise B/W	% elements closer to other centres than their own	Clustering 1 % of misclassified elements		Clustering 2 % of misclassified elements		Average number of superclusters	Average number of branches in the tree
				Hierarchical tree	Grouping	k-means (k=10)	Grouping		
	4	9.62	0.00%	1.12%	1.56%	0.76%	0.76%	3.99	5.40
	4	3.98	0.00%	2.13%	2.27%	0.91%	0.91%	4.11	6.88
Cluster of	4	2.98	0.24%	1.18%	1.49%	0.77%	1.43%	4.29	7.97
different size	4	2.15	1.38%	6.88%	7.37%	2.57%	10.94%	4	7.68
	4	1.52	3.06%	16.40%	18.64%	4.94%	14.02%	3.75	12.15
	4	1.27	6.58%	29.93%	30.40%	10.04%	6.31%	2.7	6.65
	8	9.14	0.00%	4.88%	6.31%	6.31%	3.96%	6.95	8.60
	8	4.29	0.04%	2.77%	3.98%	3.96%	3.65%	7.2	8.40
Cluster of	8	3.17	1.05%	4.62%	5.92%	3.65%	5.78%	7.05	10.77
different size	8	2.04	4.39%	12.15%	14.69%	5.78%	1.42%	6.36	14.05
	8	1.93	3.81%	40.57%	42.72%	11.42%	11.42%	4.09	12.42
	8	1.18	13.49%	54.69%	55.69%	17.38%	17.38%	2.6	12.20

Table 4. Comparison between a hierarchical and a non-hierarchical clustering, with no look ahead, computing the aesthetics-based score S_2.

	Number of original groups	Noise B/W	% elements closer to other centres than their own	Clustering 1 % of misclassified elements		Clustering 2 % of misclassified elements		Average number of superclusters	Average number of branches in the tree
				Hierarchical tree	Grouping	k-means (k=10)	Grouping		
Cluster of different size	4	9.36	0.00%	0.24%	0.25%	0.21%	0.25%	4.43	8.88
	4	4.21	0.00%	0.21%	0.22%	0.19%	0.19%	4.38	8.55
	4	3.04	0.15%	0.77%	0.95%	0.69%	0.85%	4.34	8.46
	4	2.23	1.64%	3.46%	5.22%	2.28%	4.78%	4.31	7.78
	4	1.68	3.21%	9.74%	13.73%	5.65%	13.47%	3.67	6.59
	4	1.15	6.79%	17.19%	20.71%	10.76%	20.17%	3.24	5.41
Cluster of different size	8	9.25	0.00%	7.15%	7.91%	5.77%	7.91%	6.75	9.12
	8	4.76	0.02%	4.31%	4.84%	4.01%	4.82%	7.10	9.21
	8	3.50	1.15%	4.14%	5.71%	3.33%	5.29%	7.14	9.06
	8	2.16	4.27%	12.14%	13.73%	5.54%	12.41%	6.42	7.71
	8	1.82	3.96%	29.31%	31.38%	11.77%	29.75%	4.89	6.14
	8	1.07	12.28%	42.51%	46.04%	19.14%	44.98%	3.56	6.08

Table 5. Comparison between a hierarchical and a non-hierarchical clustering, using the look ahead algorithm, searching one step ahead and computing the aesthetics-based score S_2.

	Number of original groups	Noise B/W	% elements closer to other centres than their own	Clustering 1 % of misclassified elements		Clustering 2 % of misclassified elements		Average number of superclusters	Average number of branches in the tree
				Hierarchical tree	Grouping	k-means (k=10)	Grouping		
Cluster of different size	4	9.84	0.00%	0.15%	0.25%	0.25%	0.25%	4.86	10.79
	4	4.19	0.00%	0.32%	0.64%	0.31%	0.64%	4.88	11.54
	4	3.17	0.04%	0.70%	2.14%	0.68%	2.08%	4.87	11.56
	4	2.23	1.74%	2.87%	6.31%	2.46%	6.01%	4.61	11.20
	4	1.81	5.73%	7.92%	13.87%	6.26%	13.34%	4.06	9.21
	4	1.19	6.46%	12.05%	18.91%	10.55%	18.51%	3.62	8.39
Cluster of different size	8	9.55	0.00%	3.04%	5.11%	5.11%	5.11%	7.13	11.10
	8	4.39	0.00%	3.00%	4.40%	4.37%	4.38%	7.45	10.67
	8	3.28	0.65%	3.78%	5.82%	3.40%	5.45%	7.24	10.68
	8	2.19	6.83%	6.98%	9.49%	5.84%	8.28%	6.83	9.66
	8	1.88	3.95%	19.11%	23.50%	11.34%	21.66%	5.80	9.06
	8	1.25	13.25%	32.48%	39.86%	19.74%	38.08%	4.32	8.73