

Tables and CD diagram with a statistical comparison of the AUC results of our proposal for selecting contrast patterns, using different k values over all the tested databases.

Octavio Loyola-González
octavioloyola@bioplantitas.cu

José Fco. Martínez-Trinidad Jesús Ariel Carrasco-Ochoa
fmartine@inaoep.mx ariel@inaoep.mx

Milton García-Borroto mgarciab@ceis.cujae.edu.cu

February 17, 2017

Abstract

In this document we show supplementary material for the paper entitled “A Novel Contrast Pattern Selection Method for Class Imbalance Problems” submitted to the 9th Mexican Conference on Pattern Recognition (MCPR2017).

1 Average rankings of Friedman test

Average ranks obtained by applying the Friedman procedure

Algorithm	Ranking
5	6.9309
10	5.7766
15	5.5053
20	5.4043
25	4.8936
30	5.2128
35	5.6543
40	6.0372
45	6.6277
50	6.6809
80	7.2766

Table 1: Average Rankings of the algorithms

Friedman statistic considering reduction performance (distributed according to chi-square with 10 degrees of freedom: 50.999033.

P-value computed by Friedman Test: 1.7475860780713504E-7.

2 Post hoc comparisons

2.1 CD Diagram

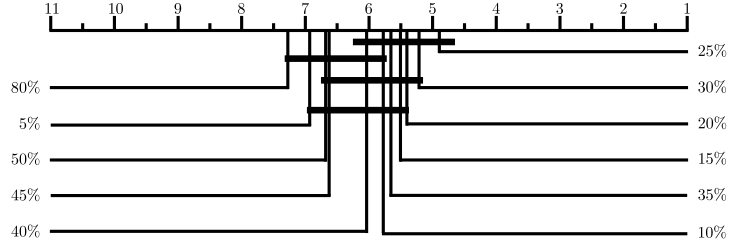


Figure 1: CD diagram with a statistical comparison (using $\alpha = 0.05$) of the AUC results of our proposal for selecting contrast patterns, using different k values over all the tested databases.

Results achieved on post hoc comparisons for $\alpha = 0.05$, $\alpha = 0.10$ and adjusted p-values.

2.2 P-values for $\alpha = 0.05$

Shaffer's procedure rejects those hypotheses that have an unadjusted p-value ≤ 0.000909 .

i	algorithms	$z = (R_0 - R_i)/SE$	p	Shaffer
55	25 vs. 80	4.925754	0.000001	0.000909
54	30 vs. 80	4.266055	0.00002	0.001111
53	5 vs. 25	4.21108	0.000025	0.001111
52	20 vs. 80	3.870236	0.000109	0.001111
51	25 vs. 50	3.694316	0.00022	0.001111
50	15 vs. 80	3.661331	0.000251	0.001111
49	25 vs. 45	3.584366	0.000338	0.001111
48	5 vs. 30	3.551381	0.000383	0.001111
47	35 vs. 80	3.353471	0.000798	0.001111
46	5 vs. 20	3.155561	0.001602	0.001111
45	10 vs. 80	3.100586	0.001931	0.001111
44	30 vs. 50	3.034617	0.002408	0.001136
43	5 vs. 15	2.946657	0.003212	0.001163
42	30 vs. 45	2.924667	0.003448	0.00119
41	20 vs. 50	2.638797	0.00832	0.00122
40	5 vs. 35	2.638797	0.00832	0.00125
39	40 vs. 80	2.561832	0.010412	0.001282
38	20 vs. 45	2.528847	0.011444	0.001316
37	15 vs. 50	2.429892	0.015103	0.001351
36	5 vs. 10	2.385912	0.017037	0.001389
35	25 vs. 40	2.363922	0.018083	0.001429
34	15 vs. 45	2.319942	0.020344	0.001471
33	35 vs. 50	2.122033	0.033835	0.001515
32	35 vs. 45	2.012083	0.044211	0.001562
31	10 vs. 50	1.869148	0.061602	0.001613
30	5 vs. 40	1.847158	0.064724	0.001667
29	10 vs. 25	1.825168	0.067976	0.001724
28	10 vs. 45	1.759198	0.078544	0.001786
27	30 vs. 40	1.704223	0.088339	0.001852
26	25 vs. 35	1.572283	0.115885	0.001923
25	45 vs. 80	1.341388	0.179794	0.002
24	40 vs. 50	1.330393	0.183389	0.002083
23	20 vs. 40	1.308404	0.190736	0.002174
22	15 vs. 25	1.264424	0.206078	0.002273
21	50 vs. 80	1.231439	0.218159	0.002381
20	40 vs. 45	1.220444	0.222297	0.0025
19	10 vs. 30	1.165469	0.243829	0.002632
18	15 vs. 40	1.099499	0.271551	0.002778
17	20 vs. 25	1.055519	0.291188	0.002941
16	30 vs. 35	0.912584	0.361461	0.003125
15	35 vs. 40	0.791639	0.428571	0.003333
14	10 vs. 20	0.769649	0.441508	0.003571
13	5 vs. 80	0.714674	0.47481	0.003846
12	25 vs. 30	0.659699	0.509447	0.004167
11	5 vs. 45	0.626714	0.530847	0.004545
10	15 vs. 30	0.604724	0.545362	0.005
9	10 vs. 15	0.560744	0.574972	0.005556
8	10 vs. 40	0.538754	0.590056	0.00625
7	20 vs. 35	0.516764	0.605321	0.007143
6	5 vs. 50	0.516764	0.605321	0.008333
5	20 vs. 30	0.39582	0.692238	0.01
4	15 vs. 35	0.30786	0.758189	0.0125
3	10 vs. 35	0.252885	0.800357	0.016667
2	15 vs. 20	0.208905	0.834523	0.025
1	45 vs. 50	0.10995	0.912449	0.05

Table 2: P-values Table for $\alpha = 0.05$

2.3 P-values for $\alpha = 0.10$

i	algorithms	$z = (R_0 - R_i)/SE$	p	Shaffer
55	25 vs. 80	4.925754	0.000001	0.001818
54	30 vs. 80	4.266055	0.00002	0.002222
53	5 vs. 25	4.21108	0.000025	0.002222
52	20 vs. 80	3.870236	0.000109	0.002222
51	25 vs. 50	3.694316	0.00022	0.002222
50	15 vs. 80	3.661331	0.000251	0.002222
49	25 vs. 45	3.584366	0.000338	0.002222
48	5 vs. 30	3.551381	0.000383	0.002222
47	35 vs. 80	3.353471	0.000798	0.002222
46	5 vs. 20	3.155561	0.001602	0.002222
45	10 vs. 80	3.100586	0.001931	0.002222
44	30 vs. 50	3.034617	0.002408	0.002703
43	5 vs. 15	2.946657	0.003212	0.002703
42	30 vs. 45	2.924667	0.003448	0.002703
41	20 vs. 50	2.638797	0.00832	0.002703
40	5 vs. 35	2.638797	0.00832	0.002703
39	40 vs. 80	2.561832	0.010412	0.002703
38	20 vs. 45	2.528847	0.011444	0.002703
37	15 vs. 50	2.429892	0.015103	0.002703
36	5 vs. 10	2.385912	0.017037	0.002778
35	25 vs. 40	2.363922	0.018083	0.002857
34	15 vs. 45	2.319942	0.020344	0.002941
33	35 vs. 50	2.122033	0.033835	0.00303
32	35 vs. 45	2.012083	0.044211	0.003125
31	10 vs. 50	1.869148	0.061602	0.003226
30	5 vs. 40	1.847158	0.064724	0.003333
29	10 vs. 25	1.825168	0.067976	0.003448
28	10 vs. 45	1.759198	0.078544	0.003571
27	30 vs. 40	1.704223	0.088339	0.003704
26	25 vs. 35	1.572283	0.115885	0.003846
25	45 vs. 80	1.341388	0.179794	0.004
24	40 vs. 50	1.330393	0.183389	0.004167
23	20 vs. 40	1.308404	0.190736	0.004348
22	15 vs. 25	1.264424	0.206078	0.004545
21	50 vs. 80	1.231439	0.218159	0.004762
20	40 vs. 45	1.220444	0.222297	0.005
19	10 vs. 30	1.165469	0.243829	0.005263
18	15 vs. 40	1.099499	0.271551	0.005556
17	20 vs. 25	1.055519	0.291188	0.005882
16	30 vs. 35	0.912584	0.361461	0.00625
15	35 vs. 40	0.791639	0.428571	0.006667
14	10 vs. 20	0.769649	0.441508	0.007143
13	5 vs. 80	0.714674	0.47481	0.007692
12	25 vs. 30	0.659699	0.509447	0.008333
11	5 vs. 45	0.626714	0.530847	0.009091
10	15 vs. 30	0.604724	0.545362	0.01
9	10 vs. 15	0.560744	0.574972	0.011111
8	10 vs. 40	0.538754	0.590056	0.0125
7	20 vs. 35	0.516764	0.605321	0.014286
6	5 vs. 50	0.516764	0.605321	0.016667
5	20 vs. 30	0.39582	0.692238	0.02
4	15 vs. 35	0.30786	0.758189	0.025
3	10 vs. 35	0.252885	0.800357	0.033333
2	15 vs. 20	0.208905	0.834523	0.05
1	45 vs. 50	0.10995	0.912449	0.1

Table 3: P-values Table for $\alpha = 0.10$

Shaffer's procedure rejects those hypotheses that have an unadjusted p-value ≤ 0.001818 .

2.4 Adjusted p-values

i	hypothesis	unadjusted p	p_{Shaf}
1	25 vs .80	0.000001	0.000046
2	30 vs .80	0.00002	0.000895
3	5 vs .25	0.000025	0.001144
4	20 vs .80	0.000109	0.004893
5	25 vs .50	0.00022	0.009922
6	15 vs .80	0.000251	0.011291
7	25 vs .45	0.000338	0.015205
8	5 vs .30	0.000383	0.017245
9	35 vs .80	0.000798	0.035912
10	5 vs .20	0.001602	0.072085
11	10 vs .80	0.001931	0.086912
12	30 vs .50	0.002408	0.089111
13	5 vs .15	0.003212	0.118855
14	30 vs .45	0.003448	0.127585
15	20 vs .50	0.00832	0.307843
16	5 vs .35	0.00832	0.307843
17	40 vs .80	0.010412	0.38525
18	20 vs .45	0.011444	0.42342
19	15 vs .50	0.015103	0.558823
20	5 vs .10	0.017037	0.613325
21	25 vs .40	0.018083	0.613325
22	15 vs .45	0.020344	0.630664
23	35 vs .50	0.033835	1.048885
24	35 vs .45	0.044211	1.370548
25	10 vs .50	0.061602	1.90967
26	5 vs .40	0.064724	1.90967
27	10 vs .25	0.067976	1.971295
28	10 vs .45	0.078544	2.199229
29	30 vs .40	0.088339	2.385164
30	25 vs .35	0.115885	2.897122
31	45 vs .80	0.179794	4.494859
32	40 vs .50	0.183389	4.494859
33	20 vs .40	0.190736	4.494859
34	15 vs .25	0.206078	4.533717
35	50 vs .80	0.218159	4.581336
36	40 vs .45	0.222297	4.581336
37	10 vs .30	0.243829	4.632757
38	15 vs .40	0.271551	4.88791
39	20 vs .25	0.291188	4.950198
40	30 vs .35	0.361461	5.783382
41	35 vs .40	0.428571	6.428567
42	10 vs .20	0.441508	6.428567
43	5 vs .80	0.47481	6.428567
44	25 vs .30	0.509447	6.428567
45	5 vs .45	0.530847	6.428567
46	15 vs .30	0.545362	6.428567
47	10 vs .15	0.574972	6.428567
48	10 vs .40	0.590056	6.428567
49	20 vs .35	0.605321	6.428567
50	5 vs .50	0.605321	6.428567
51	20 vs .30	0.692238	6.428567
52	15 vs .35	0.758189	6.428567
53	10 vs .35	0.800357	6.428567
54	15 vs .20	0.834523	6.428567
55	45 vs .50	0.912449	6.428567

Table 4: Adjusted p -values