Maximizing Area Under the Receiver Operating Characteristic Curve for Biomarker Combination

Yuxuan Chen and Yijian Huang

Department of Biostatistics and Bioinformatics, Emory University

Supplementary Materials

We present the simulation results for a setting of 8 biomarkers. Five non-informative biomarkers were added to each of the set-ups considered in Section 3.1. These five biomarkers were independent of each other and of other biomarkers and each followed the standard normal distribution regardless of the disease status.

The computational performance is provided in Table S1. Due to the inclusion of the additional biomarkers, the maximized empirical AUC became substantially larger, especially with smaller sample sizes. This upward bias shrinked to zero as the sample size increased. The proposed method provides the largest empirical AUC in all settings with generally shorter computation time.

The statistical performance is shown in Table S2. To illustrate the estimation performance for noninformative biomarkers, the bias and standard deviation of the coefficient estimation for the fourth biomarker are included. As expected, the biases of the first three coefficients were larger compared to the three-biomarker scenario. The bias of the coefficient estimate for the noninformative biomarker is close to zero; however, its standard deviation is larger than those of the other coefficient estimates. This is not surprising since there is no true association between the noninformative biomarker and the outcome. The relative performance is similar to what was observed in the three-biomarker settings. Overall, our proposed method achieves coefficient estimation with similar bias and standard deviation as other AUC maximization methods.

Table S1: Simulation study results on the computational performance of coefficient estimation

n	100		200		500		2000	
Method	A	T	A	T	A	T	A	T
Scenario I								
PROP	90.26	0.10	88.40	0.21	87.16	0.90	86.62	11.55
SAUC	90.07	0.34	88.30	0.30	87.10	1.05	86.61	11.66
GAUC	90.09	1.66	88.27	1.42	87.08	1.28	86.61	7.00
RAUC	89.55	1.76	88.04	20.78	-	-	-	-
SMRCE	89.25	0.04	87.91	0.13	86.98	0.53	86.60	12.67
SANN	89.28	0.10	87.73	0.19	87.01	0.79	86.62	9.75
Scenario II								
PROP	90.14	0.19	88.14	0.28	86.97	0.77	86.36	5.39
SAUC	90.06	0.34	88.02	0.33	86.89	1.55	86.33	13.61
GAUC	90.09	1.96	87.99	1.90	86.87	3.18	86.33	9.89
RAUC	89.54	1.87	87.76	21.25	-	-	-	-
SMRCE	89.24	0.04	87.62	0.13	86.77	0.84	86.32	12.21
SANN	88.94	0.10	87.49	0.19	86.75	0.77	86.27	9.72

A: Optimized empirical AUC ($\times 100$); T: computation time (in seconds); PROP: proposed method; SAUC: Sigmoid-smoothed AUC; GAUC: Gaussian-smoothed AUC; RAUC: ramp AUC; SMRCE: smoothed maximum rank correlation estimator; SANN: simulated annealing. RAUC was not performed for larger sample sizes due to lengthy running time.

Table S2: Simulation study results on coefficient estimation

	Scenario I								Scenario II							
	β_1		β_2		β_3		β_4		β_1		β_2		β_3		β_4	
Method	B	SD	B	SD	B	SD	B	SD	B	SD	B	SD	B	SD	B	SD
								$n=100$								
PROP	-69	151	-75	150	-61	149	-10	173	-135	143	33	143	-68	139	-0	179
LR	-67	145	-67	140	-53	143	-9	171	-128	135	59	134	-81	141	-1	174
SAUC	-72	152	-75	153	-60	150	-10	176	-137	143	35	144	-71	141	2	182
GAUC	-76	156	-80	157	-66	154	-11	183	-142	148	28	148	-74	143	1	185
RAUC	-73	151	-73	151	-61	152	-11	176	-131	144	35	143	-80	143	0	182
SMRCE	-67	135	-70	134	-55	132	-10	177	-131	124	36	128	-67	139	-1	180
SANN	-67	145	-57	137	-62	142	-3	173	-131	140	57	143	-85	148	3	177
								$n=200$								
PROP	-32	106	-37	104	-33	110	-2	124	-89	107	48	110	-37	101	-0	139
LR	-29	101	-34	99	-28	104	-0	116	-88	102	87	106	-61	103	-0	132
SAUC	-34	107	-37	105	-32	110	-3	124	-88	107	49	111	-41	101	-1	139
GAUC	-35	110	-40	108	-35	114	-2	129	-91	109	43	112	-39	102	-1	142
RAUC	-32	106	-38	104	-33	111	-2	125	-85	108	48	110	-44	105	-1	139
SMRCE	-30	92	-34	91	-28	96	-2	122	-90	94	57	94	-37	101	-1	139
SANN	-30	103	-34	102	-31	99	-7	118	-91	101	87	102	-56	101	-5	128
								$n=500$								
PROP	-13	65	-10	67	-19	67	-6	85	-66	68	61	69	-7	62	1	89
LR	-13	63	-8	65	-17	63	-4	79	-66	64	102	65	-36	65	-0	83
SAUC	-14	67	-11	68	-19	68	-6	87	-67	69	60	70	-7	63	2	90
GAUC	-15	68	-10	69	-19	69	-5	87	-65	70	57	71	-7	63	2	92
SMRCE	-12	56	-9	58	-16	57	-5	83	-68	61	67	58	-6	63	0	89
SANN	-10	63	-14	63	-13	62	-1	78	-65	60	101	65	-36	63	1	85
								$n=2000$								
PROP	-3	32	-4	34	-4	34	0	41	-53	34	63	36	11	30	1	44
LR	-2	30	-4	32	-3	32	1	39	-57	32	109	34	-22	32	2	41
SAUC	-2	33	-4	34	-4	34	0	41	-53	34	63	36	11	30	1	45
GAUC	-2	33	-4	34	-4	34	-0	41	-52	34	62	36	11	30	1	44
SMRCE	-2	29	-4	30	-4	30	0	40	-53	30	64	32	11	30	1	44
SANN	-2	31	-3	32	-5	33	1	40	-56	31	109	33	-22	32	2	42

B: bias ($\times 1000$); SD: standard deviation ($\times 1000$).

PROP: proposed method; LR: logistic regression; SAUC: Sigmoid-smoothed AUC; GAUC: Gaussian-smoothed AUC; RAUC: ramp AUC; SMRCE: smoothed maximum rank correlation estimator; SANN: simulated annealing. RAUC was not performed for larger sample sizes due to lengthy running time.