unpartnered_partnered_timing_analysis

Fiona Yeung

May 1, 2018

R Markdown

This timing analysis shows that the number of characteristics group is the main driver of the run time. I believe the computational overhead from matrix processing also increases with larger network size. Hence the run time also increases with the network size.

The code that generated this result is 100% R code with no parallelization. This timing analysis was done on my laptop with 8th Gen Intel Core i7-8550U processor (1.80GHz 8MB) and 16GB DDR4 2400 MHz memory.

network size	# coeff	# groups	run time	# paired indiv.	# unpaired indiv.
14527	14	8	2.89	1126	13401
28408	14	8	3.27	2032	26376
48447	14	8	3.27	3552	44895
65963	14	8	3.58	4538	61425
131926	14	8	4.37	9076	122850
263852	14	8	6.20	18152	245700
527704	14	8	9.92	36304	491400

network size	# coeff	# groups	run time	# paired indiv.	# unpaired indiv.
14527	12	32	70.72	1126	13401
28408	12	32	76.57	2032	26376
48447	12	32	70.07	3552	44895
65963	12	32	72.59	4538	61425
131926	12	32	78.53	9076	122850
263852	12	32	83.92	18152	245700
527704	12	32	90.54	36304	491400

network size	# coeff	# groups	run time
14527	10	8	1.39
14527	14	8	2.89
14527	4	32	29.49
14527	6	32	44.98
14527	12	32	70.72

network size	# coeff	# groups	run time
527704	10	8	8.36
527704	14	8	9.92
527704	4	32	49.49
527704	6	32	60.22
527704	12	32	90.54

Raw numbers from all runs:

network size	# coeff	# groups	elapsed time	# paired indiv.	# unpaired indiv.	user time	system time	formula
14527	10	8	1.39	1126	13401	1.39	0.00	1
28408	10	8	2.11	2032	26376	2.11	0.00	1
48447	10	8	2.03	3552	44895	2.03	0.00	1
65963	10	8	1.94	4538	61425	1.90	0.00	1
131926	10	8	2.67	9076	122850	2.66	0.02	1
263852	10	8	4.89	18152	245700	4.86	0.01	1
14527	6	32	44.98	1126	13401	44.90	0.02	2
28408	6	32	48.85	2032	26376	48.49	0.05	2
48447	6	32	44.39	3552	44895	44.33	0.00	2
65963	6	32	45.27	4538	61425	45.10	0.00	2
131926	6	32	45.72	9076	122850	45.62	0.04	2
263852	6	32	52.50	18152	245700	52.28	0.08	2
14527	12	32	70.72	1126	13401	70.58	0.00	3
28408	12	32	76.57	2032	26376	76.50	0.00	3
48447	12	32	70.07	3552	44895	69.95	0.00	3
65963	12	32	72.59	4538	61425	72.50	0.00	3
131926	12	32	78.53	9076	122850	78.44	0.03	3
263852	12	32	83.92	18152	245700	83.78	0.05	3
14527	14	8	2.89	1126	13401	2.89	0.02	4
28408	14	8	3.27	2032	26376	3.27	0.00	4
48447	14	8	3.27	3552	44895	3.26	0.00	4
65963	14	8	3.58	4538	61425	3.58	0.00	4
131926	14	8	4.37	9076	122850	4.37	0.00	4
263852	14	8	6.20	18152	245700	6.16	0.03	4
14527	4	32	29.49	1126	13401	29.43	0.00	5
28408	4	32	31.15	2032	26376	31.11	0.00	5
48447	4	32	32.05	3552	44895	32.03	0.00	5
65963	4	32	32.52	4538	61425	32.47	0.01	5
131926	4	32	35.15	9076	122850	35.13	0.00	5
263852	4	32	42.08	18152	245700	42.01	0.03	5
527704	10	8	8.36	36304	491400	8.30	0.02	1
527704	6	32	60.22	36304	491400	58.92	0.16	2
527704	12	32	90.54	36304	491400	89.59	0.17	3
527704	14	8	9.92	36304	491400	9.90	0.00	4
527704	4	32	49.49	36304	491400	49.15	0.09	5

Example of a linear utility function (the one that took the longest in our timing analysis above): ff4 = ~ b1nodematch("educlevel_t") + b2nodematch("educlevel_t") + b1absdiff("educlevel_t",1) + b2absdiff("educlevel_t",1) + b1greaterthan("educlevel_t") + b2greaterthan("educlevel_t")

$$
\begin{aligned}
U(x_i, z_j|\theta_{Ui}) =& \theta_{U0} + \theta_{U1} I(< HS_i, < HS_j) + \theta_{U2} I(HS_i, HS_j) + \\
& \theta_{U3} I(SomeCollege_i, SomeCollege_j) + \theta_{U4} I(BA+_i, BA+_j) + \\
& \theta_{U5} I(< HS_i, HS_j \mid\mid HS_i, < HS_j \mid\mid HS_i, SomeCollege_j \mid\mid \\
& SomeCollege_i, HS_j \mid\mid SomeCollege_i, BA+_j \mid\mid BA+_i, SomeCollege_j) + \\
& \theta_{U6} I(< HS_i, HS_j \mid\mid < HS_i, SomeCollege_j \mid\mid < HS_i, BA+_j \mid\mid \\
& HS_i, SomeCollege_j \mid\mid HS_i, BA+_j \mid\mid SomeCollege_i, BA+_j)
\end{aligned}
$$

	male: <HS	male: HS	male: SomeCollege	male: BA+	single female
female: <HS	1168	1248	560	64	31560
female: HS	1488	5424	2896	912	74480
female: SomeCollege	816	3888	5456	2208	95896
female: BA+	128	1280	2512	6256	53808
single male	33864	81568	79232	40992	NA