HCP data analysis

Chanwoo Lee January 15, 2021

The HCP dataset records the structural connectivity patterns among 68 brain regions for 114 individuals. The resulting dataset is an order-3 tensor $\mathcal{Y} \in \{0,1\}^{68\times 68\times 114}$, where the entries encode the presence or absence of fiber connections between those 68 brain regions.

We first investigate the prediction accuracy of tensor completion. We diversify the missing rates as $20\%, 33\%$, $50\%, 67\%$ controlling the number of the training points and check misclassification rate(MCR) on testing which is binary version of MAE. Figure1 shows the MCR across different ranks and missing rates when $H = 10$. We find that our method substantially outperforms the classical low-rank method in most cases. Furthermore, our method performs much better especially when the missing rate is getting high. It turns out that increment of rank has little effect on the performance. Overall, the rank $r = 10$ is enough for good performance and increment of rank does not improve the performance when missing rate is greater than 33%.

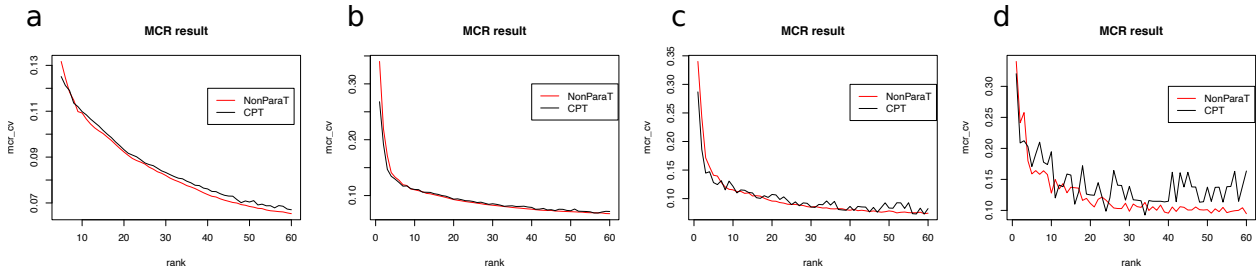

Figure 1: Misclassification rate across different missing rates. a: 20% missing rate, b: 33% missing rate, c: 50% missing rate, d: 67% missing rate.

The following tables focus on the prediction accuracy when rank is $9, 12, 15$ according to different missing rates $(20\%, 33\%, 50\%, 67\%)$. When the rank is 3 or 6, CPT methods performs better than ours. However, except those two ranks, our method outperforms substantially in all cases with much smaller standard errors.

rank	9				12				15			
Missing rate	20%	33%	50%	67%	20%	33%	50%	67%	20%	33%	50%	67%
NonparaT (Ours)	0.109 (0)	0.111 (4)	0.118 (16)	0.157 (73)	0.104 (1)	0.106 (3)	0.111 (9)	0.140 (35)	0.100 (1)	0.101 (3)	0.106 (10)	0.133 (97)
Low-rank CPT	0.111 (3)	0.113 (7)	0.122 (57)	0.154 (117)	0.106 (4)	0.108 (7)	0.112 (63)	0.150 (123)	0.101 (3)	0.103 (6)	0.107 (37)	0.151 (176)

Table 1: Prediction accuracy measured in MCR in the HCP data analysis. The reported MCRs are averaged over five runs of cross-validation with standard errors in parenthesis ($\times 10^{-3}$). We set $H = 10$ in the analysis.

For future plan, I am going to estimate the signal tensor $\hat{\Theta}$ from the full tensor $\mathcal{Y}$, $r = 10$, and $H = 10 or 20$. We can find out the structural connectivity patterns of brain nodes by checking the top 4 tensor components across the 68 brain regions. I think that plotting the edges with high loadings based on the tensor components can reveal some interesting spatial patterns.