Minimax Structured Neural Tangent Kernel in Estimating Average Treatment Effect Confounded by Image Covariate

EMORY
Undergraduate Research Programs
Pathways Center

Zijian (Michael) Wang
Department of Quantitative Theory and Methods

Introduction & Motivation

Problem: ATE estimation is biased with high-dimensional image covariates. Example application: lung pneumonia imaging.

Solution: Propose a minimax framework using Neural Tangent Kernel to reduce bias in IPW estimation.

Empirical Testing: Using Semi-Synthetic Data Generated from lung imaging to test proposed methods: 1) Brightness-based Framework; 2) Label + Brightness; 3) Filtered Image Features.

Methods/Approach

We could write the imbalance as a worst-case scenario over a class of functions expressing conditional means as:

$$\text{imbalance}_{\mathcal{F}}(\gamma) := \max_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^{n} \gamma_i f(W_i, X_i) - \frac{1}{n} \sum_{i=1}^{n} [f(1, X_i) - f(0, X_i)] \right|$$

Then the ideal balancing weights could be minimized to be:

$$\gamma^* = \arg\min_{\gamma} \left\{ \text{imbalance}_{\mathcal{F}}(\gamma) + \frac{\sigma^2}{n^2} \|\gamma\|^2 \right\}$$

With the freedom in selecting those functions in high-dimensional settings, we propose Neural Tangent Kernel (NTK) based approach for function classes in the Reproducing Kernel Hilbert Space (RKHS), expressed as Kernel functions.

$$K_{NTK}(x, x') = \nabla_\theta f(x)^\top \nabla_\theta f(x')$$

The estimated balancing weights becomes:

$$\hat{\gamma} = (K_{\text{NTK}} + \lambda(B) \cdot I)^{-1} K_{\text{diff}}$$

Results

Method	$\hat{\tau}_0$	**Sample** σ	**Coverage**	$E_n[\hat{\tau}_i]$
Truth	1	NA	NA	NA
IPW I	0.863	0.197	0.93	1.043
IPW II	0.898	0.283	0.99	1.065
IPW III	0.765	0.072	0.71	0.893
IPW w/ Linear Mini-Max	0.982	0.017	0.94	0.997
IPW w/ RBF Mini-Max	1.001	0.007	0.93	1.001
IPW w/ NTK Mini-Max	1.070	0.022	0.32	1.052
AIPW w/ NTK Mini-Max	1.002	0.021	0.93	0.999

Table 8.1: Semi-Synthetic Framework 1

Method	$\hat{\tau}_0$	**Sample** σ	**Coverage**	$E_n[\hat{\tau}_i]$
Truth	0.227	NA	NA	NA
IPW I	0.362	0.137	0.95	0.254
IPW II	0.337	0.116	0.96	0.242
IPW III	0.303	0.057	0.68	0.308
IPW w/ Linear Mini-Max	0.273	0.062	0.96	0.230
IPW w/ RBF Mini-Max	0.292	0.074	0.94	0.237
IPW w/ NTK Mini-Max	0.284	0.078	0.93	0.263
AIPW w/ NTK Mini-Max	0.323	0.061	0.43	0.356

Table 8.2: Semi-Synthetic Framework 2

Method	$\hat{\tau}_0$	**Sample** σ	**Coverage**	$E_n[\hat{\tau}_i]$
Truth	1	NA	NA	NA
IPW I	1.006	0.195	0.94	1.008
IPW II	0.997	0.091	0.95	1.023
IPW III	0.812	0.073	0.44	0.846
IPW w/ Linear Mini-Max	0.951	0.051	0.97	0.959
IPW w/ RBF Mini-Max	0.985	0.022	0.95	0.994
IPW w/ NTK Mini-Max	1.072	0.041	0.05	1.149
AIPW w/ NTK Mini-Max	1.123	0.004	0	1.121

Table 8.3: Semi-Synthetic Framework 3

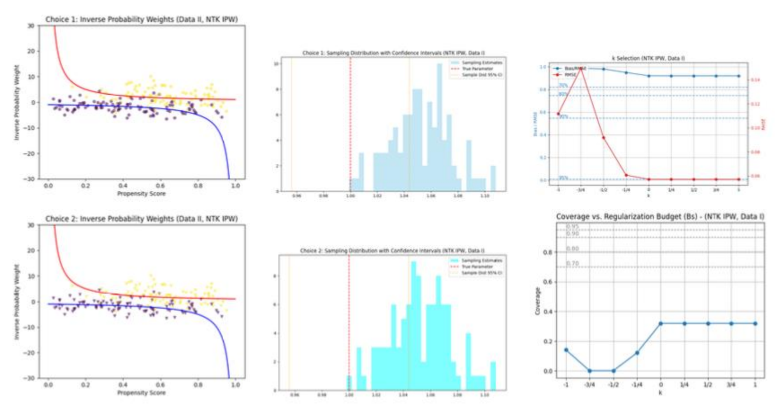

Figure A.4: Semi-Synthetic Framework 1 - IPW w/ NTK Minimax

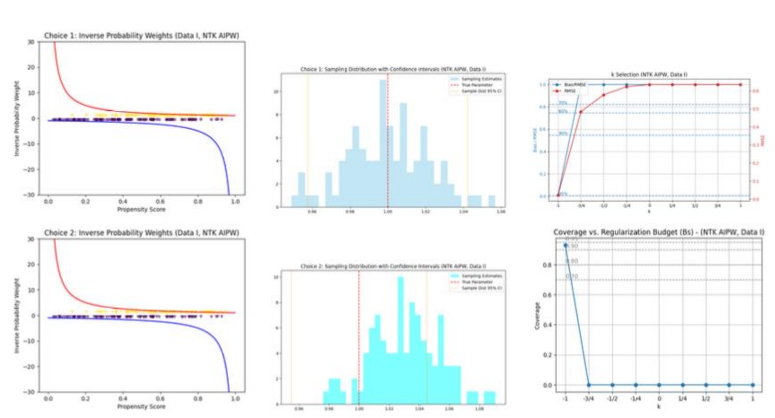

Figure A.5: Semi-Synthetic Framework 1 - AIPW w/ NTK Minimax

Estimators we've tested: 1) IPW with true and estimated propensity scores (IPW I, II, III); 2) IPW w/ Linear Minimax (linear conditional mean functions); 3) IPW w/ RBF Minimax (RBF Kernel functioned conditional mean); 4) IPW & AIPW w/ NTK Minimax (NTK functioned conditional mean).

Our empirical evaluation across three semi-synthetic frameworks reveals that oracle estimators—such as Linear and RBF-based Minimax IPW—consistently achieve low bias and high coverage, validating their effectiveness under known covariate structured function expressions. While non-oracle estimators like NTK-based IPW and AIPW offer a flexible, model-free approach, their performance is more variable and sensitive to regularization. Unfortunately, NTK-AIPW suffered from under-coverage and instability in complex image-confounded data, likely due to insufficient fine tuning in the neural network component as well as parameter tuning.

Conclusion/Future Directions

	$\hat{\tau}_0$	**Sample** σ	**Coverage**	$E_n[\hat{\tau}_i]$	$\lambda(B)$	B
Truth	1					
Choice 1	1.070	0.022	0.32	1.052	2.5e-5	1
Choice 2	1.070	0.022	0.32	1.052	6.9e-13	6000

Table B.3: Parameter Tuning: Data 1, IPW w/ NTK Minimax

	$\hat{\tau}_0$	**Sample** σ	**Coverage**	$E_n[\hat{\tau}_i]$	$\lambda(B)$	B
Truth	1					
Choice 1	1.002	0.021	0.93	0.999	3600	0.00016
Choice 2	1.034	0.022	0.73	1.030	3000	0.00146

Table B.4: Parameter Tuning: Data 1, AIPW w/ NTK Minimax

While traditional oracle methods perform well as oracle estimators when covariate structure is partially known, NTK-based estimators offer a promising yet currently limited solution when such knowledge is unavailable. In our implementation, the NTK based approaches admittedly doesn't perform as well as expected potentially due to their sensitivities to neural network predictions and insufficient parameter tuning.

Future work will focus on improving the neural network component—potentially incorporating deeper architectures or transfer learning—to enhance the expressiveness of the NTK. Also, better parameter tuning mechanisms would be deployed to improve the performance of our Minimax structured estimators, particularly the NTK estimators. Additionally, expanding this framework to real-world medical imaging datasets.

Acknowledgements

My thank to my thesis committee members: Dr. David Hirshberg, Dr. Kevin McAlister, and Dr. Abhishek Ananth for their support.