# Deep LTMLE

Toru Shirakawa, Yi Li, Yulun Wu, Sky Qiu, Yuxuan Li, Mingduo Zhao, Hiroyasu Iso , Mark van der Laan
Osaka University Graduate School of Medicine; University of California, Berkeley

## Longitudinal Targeted Minimum Loss-based Estimation with Temporal-Difference Heterogeneous Transformer

**Deep LTMLE** estimates the mean counter factual outcome under dynamic interventions from complex longitudinal observational data.

built upon LTMLE (van der Laan and Gruber, 2011: sequential regression + targeting + ensemble ML) and DeepACE (Frauen et al, 2023: stochastic LTMLE with RNN)

improved LTMLE which cannot pool information across time and was not scalable due to its sequential algorithm

improved DeepACE which did not provide uncertainty measure and was not scalable due to its sequential algorithm

### Longitudinal Data

$$W \to L_1 \to^{g_1}_{\pi_1} A_1 \to Y_1 \to \cdots \to L_\tau \to^{g_\tau}_{\pi_\tau} A_\tau \to Y_\tau \qquad g_t(\cdot \mid pa(A_t)), \pi_t(\cdot \mid pa(A_t)) \in \mathcal{P}(\mathcal{A}_t)$$

### Target parameter (identified through g-formula)

$$\psi(P) = EY^g = E_g Y \qquad g \ll \pi, Y^g \perp A_t \mid pa(A_t)$$

### Estimator (Sequential Regression with a Summed Loss)

Similar to fitted Q-iteration

$$\hat{Y}_\tau \to (reg) \to \hat{Q}_\tau(A_\tau, pa(A_\tau)) \to \int \hat{Q}_t dg_\tau \to \hat{V}_\tau \to (reg) \to \hat{Q}_{\tau-1}$$

$$\cdots \to \hat{Q}_1 \to \int \hat{Q}_1 dg_1 \to \hat{V}_1 \to \hat{\psi} = P_n \hat{V}_1$$

### von Mises Expansion (Infinite Dimensional Taylor Expansion)

$$\psi(P) - \psi(P_0) = -\int D^\star dP_0 + R_2(P, P_0)$$

### Targeting (Debiasing)

Canonical gradient $D^\star = \nabla\psi$ (the Riesz representor of the first variation of $\psi$):

$\partial_\varepsilon|_0 \psi(P_\varepsilon) = \langle D^\star(P), h \rangle_P$ for any path $(P_\varepsilon)_\varepsilon$ on the model $\mathcal{M}$ with a score $h \in \mathbb{L}_0^2(P)$

For our g-functional, the canonical gradient is given as follows (van der Laan and Gruber, 2011)

$$D^\star(Q,G)(O) = V_1 - \psi_0 + \sum_t^\tau I_t(G)(V_t - Q_t) \qquad G_t = \frac{dg_t}{dv_t}(O), I_t = \frac{dg_{1:t}}{d\pi_{1:t}}(O)$$

Submodel

$$\operatorname{logit} Q_{t,\varepsilon} = \operatorname{logit} Q_t + \varepsilon$$

Loss Function

$$\mathcal{L}(Q, V) = \sum_{t=1}^{\tau} I_t(G) \mathcal{L}_{bce}(Q_t, V_t)$$

TMLE $\quad \varepsilon^\star = \arg\min_\varepsilon \mathcal{L}(\hat{Q}_\varepsilon, \hat{V}_{\varepsilon^\star}) \qquad \hat{Q}^\star = \hat{Q}_{\varepsilon^\star} \qquad \hat{\psi}^\star = \psi(\hat{Q}^\star)$

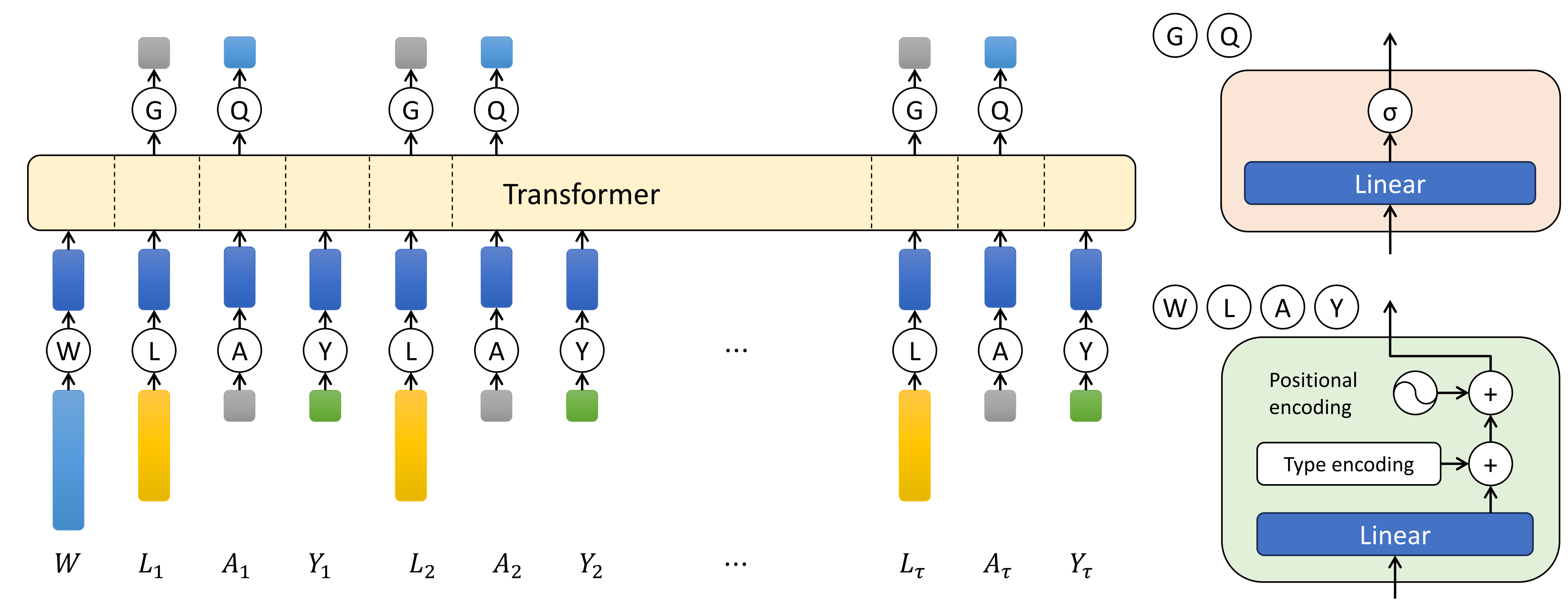
### Inference (Estimator of Asymptotic Variance)

$$\hat{\sigma}_n = P_n D^\star(\hat{Q}^\star, \hat{G})^2$$

Paper

Python Package

### Architecture of Temporal-Difference Heterogeneous Transformer




### Simulation Results

| Model | Bias $\tau=10$ | Bias $\tau=20$ | Bias $\tau=30$ | RMSE $\tau=10$ | RMSE $\tau=20$ | RMSE $\tau=30$ | Coverage $\tau=10$ | Coverage $\tau=20$ | Coverage $\tau=30$ | Mean $\hat{\sigma}_n$ $\tau=10$ | Mean $\hat{\sigma}_n$ $\tau=20$ | Mean $\hat{\sigma}_n$ $\tau=30$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| LTMLE (GLM) | 0.0230 | 0.0766 | 0.1344 | 0.0265 | 0.0796 | 0.1381 | 1.00 | 1.00 | 1.00 | 0.43 | 0.69 | 0.76 |
| LTMLE (SL) | 0.0144 | 0.0297 | 0.0477 | 0.0185 | 0.0344 | 0.0545 | 1.00 | 1.00 | 1.00 | 0.31 | 0.40 | 0.45 |
| DeepACE | 0.0055 | -0.0211 | -0.0714 | 0.0397 | 0.0672 | 0.0962 | 1.00 | 1.00 | 1.00 | 0.69 | 0.76 | 0.73 |
| Deep LTMLE | 0.0181 | 0.0292 | 0.0503 | 0.0263 | 0.0332 | 0.0533 | 0.99 | 0.96 | 0.75 | 0.17 | 0.10 | 0.06 |
| Deep LTMLE† | 0.0319 | 0.0327 | 0.0535 | 0.0571 | 0.0379 | 0.0571 | 0.99 | 0.97 | 0.78 | 0.19 | 0.10 | 0.07 |
| Deep LTMLE⋆ | 0.0156 | 0.0307 | 0.0496 | 0.0218 | 0.0344 | 0.0541 | 0.99 | 0.95 | 0.78 | 0.17 | 0.09 | 0.07 |

### Causal Effects of Blood Pressure Management Strategies after 30 years of follow-up