Abstract:
TiDE is a Multi-layer Perceptron (MLP) based encoder-decoder model that uses the simplicity and speed of linear models while also handling covariates and non-linear dependencies without temporal self-attention, offering an attractive accuracy-efficiency trade-off. TiDE's own ablations leave open a different question: whether attention helps when applied at the point where TiDE combines its input segments-the target look-back, the past and future covariates, and the static attributes- which the architecture currently fuses by flat concatenation. We introduce GA-TiDE, which modifies TiDE in two places. First, every residual block acquires a sigmoid gate on its nonlinear branch, leaving the activation, the dropout placement, and the linear skip untouched, so that the modification is a single element-wise factor. Second, each flattened input segment is projected to a common width, treated as one token, and passed through a single multi-head self-attention layer before entering the encoder. Additionally, we uncover and fix a gradient degeneracy in vanilla TiDE: when use_layer_norm=True is combined with a univariate target, the LayerNorm(1) operator analytically zeroes the entire encoder gradient, silently degenerating the model to a near-linear persistence predictor. Our fix conditionally skips LayerNorm when the output width is one, restoring full gradient flow. We provide diagnostics showing the degeneracy is analytic - identically zero in both float32 and float64. Across standard long-horizon forecasting benchmarks, GA-TiDE outperforms vanilla TiDE on several datasets while adding modest parameter overhead, though gains are not uniform across all horizons and datasets, and the added attention and gating machinery increase training memory and runtime relative to the baseline.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.