NANJING UNIVERSITY

Feature-aware Modulation for Learning from Temporal Tabular Data

Hao-Run Cai Han-Jia Ye

School of Artificial Intelligence, Nanjing University

NEURAL INFORMATION PROCESSING SYSTEMS

Summary

- Challenge: Static models fail under temporal shifts because feature semantics evolve.
- Insight: Distributional statistics (mean, std, skewness) act as effective proxies.
- Method: Feature-aware modulation conditioned on temporal context.

Learning from Temporal Tabular Data

- Tabular data are often collected over time, resulting in temporal shifts at each time point.
- Relying on the *i.i.d.* assumption, static models fail to capture the dynamic nature of temporal shifts.
- Concept shift: "High Income" implies different values in 2010s *vs.* 2020s.

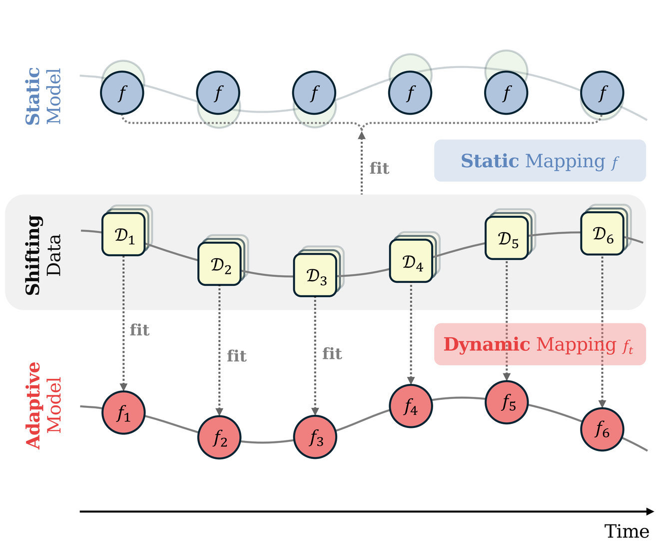

Figure 1. Static model f vs. adaptive model f_t.

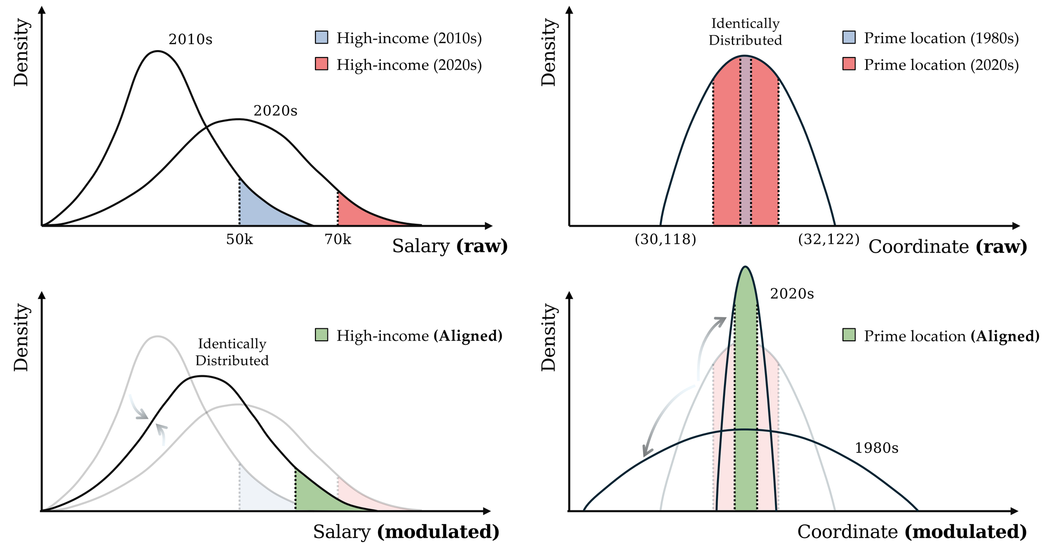

Figure 2. Modulating feature distributions for concept consistency.

Aligning Feature Concept under Temporal Shift

We seek *semantic invariance*:

$$\phi(\boldsymbol{x};\theta_t) = \phi(\boldsymbol{x}';\theta_{t'}), \quad \text{if } \boldsymbol{x} \sim \mathcal{D}_t,\ \boldsymbol{x}' \sim \mathcal{D}_{t'}, \text{ and } \boldsymbol{x}, \boldsymbol{x}' \text{ are semantically equivalent.}$$

Figure 3 illustrates the temporal distributions of selected features from **real-world datasets** in TabReD benchmark (Rubachev et al., 2025). Distributional statistics *e.g.* mean, std, skewness can effectively characterize the majority of temporal distribution shifts observed.

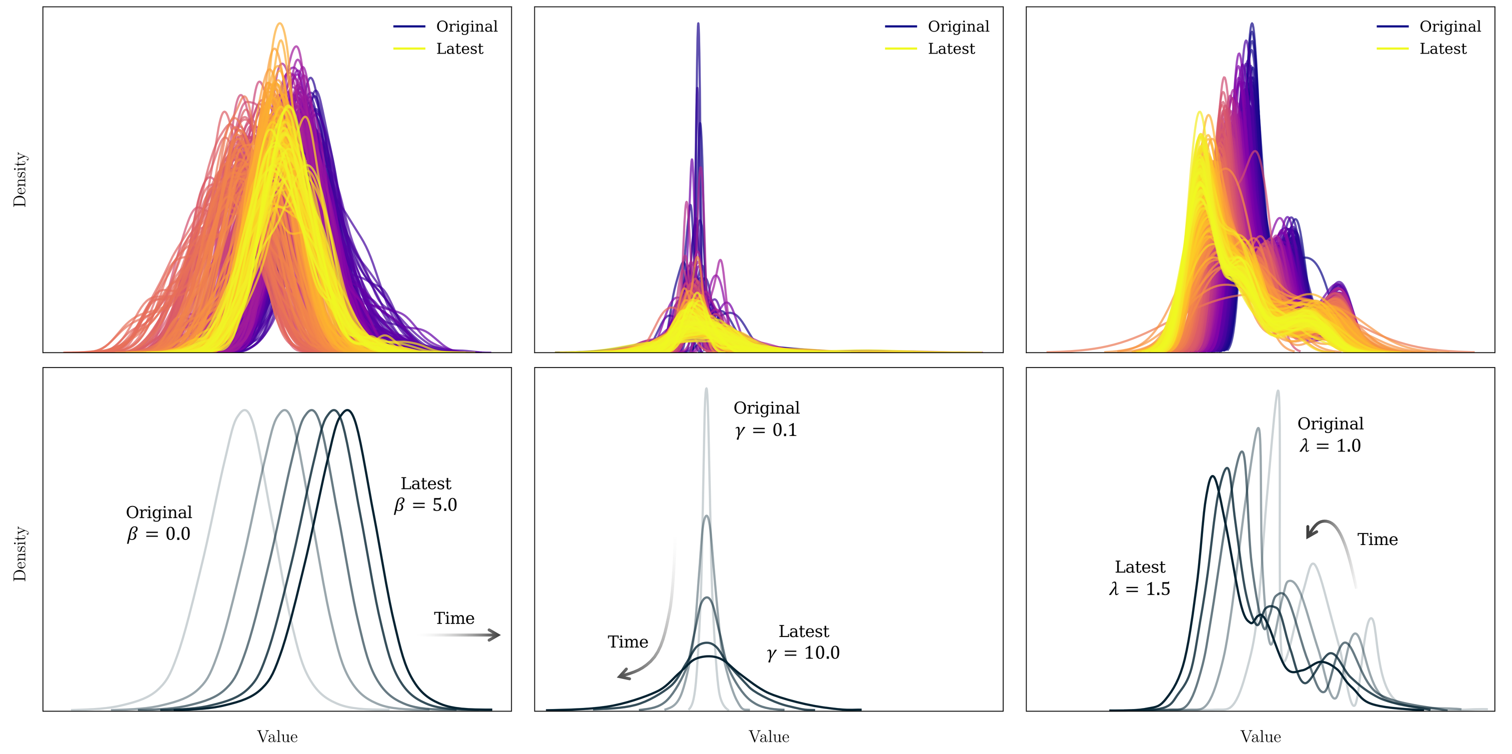

Figure 3. Empirical feature distributions over time, and schematic illustration of learnable transformations applied to feature distributions.

Feature-aware Temporal Modulation

Modulation Scheme

Concretely, for each feature $\boldsymbol{x} \in \mathbb{R}^m$ and $t \in \mathbb{R}$, the modulator produces $\gamma(\psi(t)), \beta(\psi(t)), \lambda(\psi(t)) \in \mathbb{R}^m$, which correspond to **scale**, **bias**, and a **nonlinear transformation coefficient**, respectively. We define the temporal modulation function as follows:

$$\tilde{\boldsymbol{x}}_i = \gamma_i(\psi(t)) \cdot \mathrm{YJ}(\boldsymbol{x}_i; \lambda_i(\psi(t))) + \beta_i(\psi(t)), \tag{1}$$

where $\mathrm{YJ}(\boldsymbol{x}_i; \lambda_i)$ denotes the Yeo-Johnson transformation, a non-linear transformation handling asymmetry.

Network Architecture

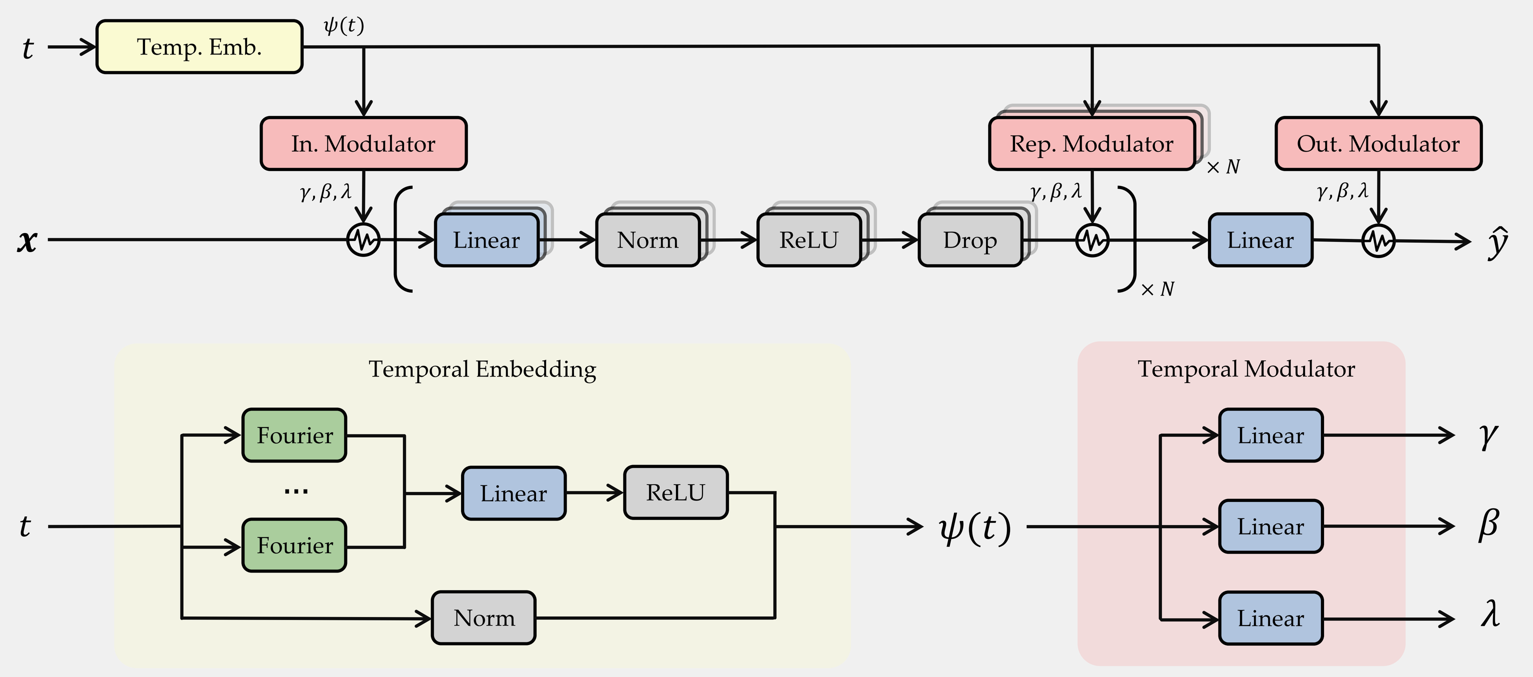

Figure 4. Overview of our temporal modulation framework. Parameters γ, β, λ are used for modulation defined in eq. (1).

Pilot Study

- Static View: Without temporal context, the task appears chaotic and completely **non-separable**.
- Temporal Dynamics: Integrating timestamps allows the model to **adaptively adjust** decision boundaries.
- Modulated View: In our aligned representation, the decision boundary becomes **stable and unified**.

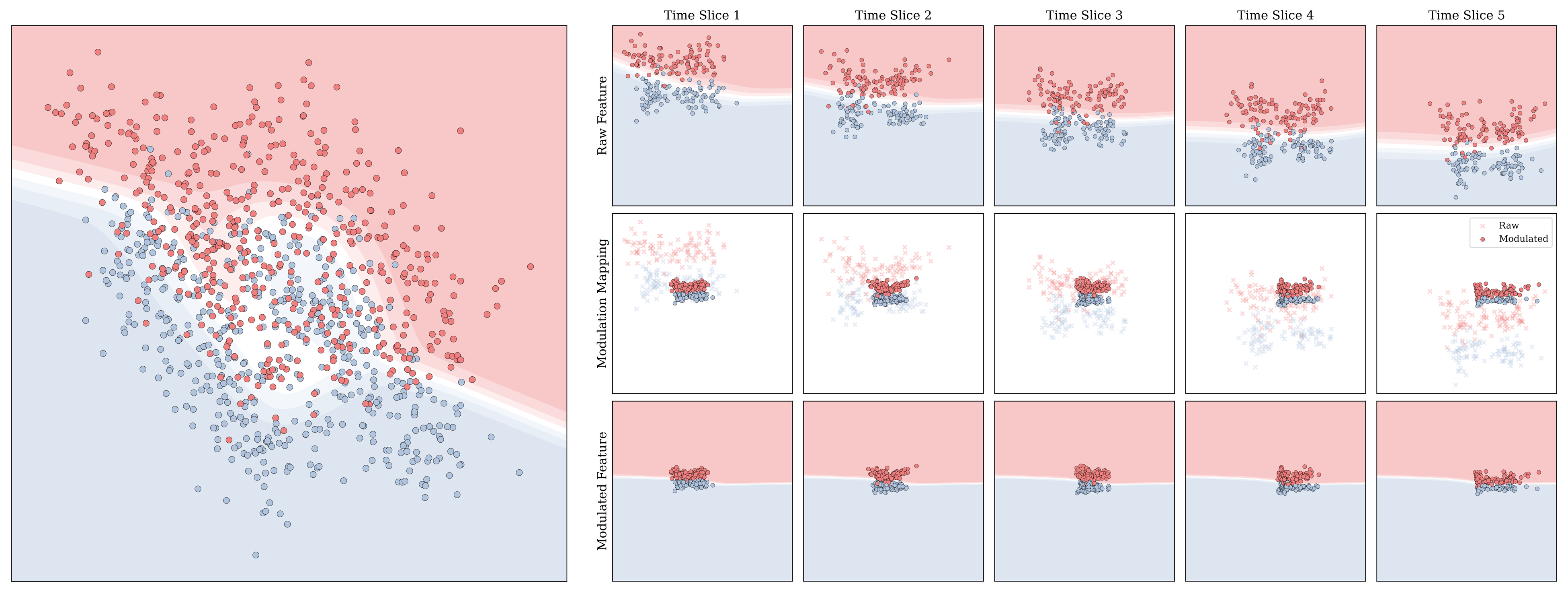

Figure 5. Pilot study on aligning feature semantics.

Experiments

We conduct experiments on the TabReD benchmark proposed by Rubachev et al. (2025), using the refined training protocol introduced by Cai & Ye (2025).

	Methods	HI↑	EO↑	HD↑	SH↓	CT↓	DE↓	MR↓	WE↓	Rank
Static Methods	**Classical Baselines**									
	Linear	0.9388	0.5944	0.8231	0.2435	0.4864	0.5596	0.1680	1.7464	18.250
	XGBoost	0.9625	0.6199	**0.8644**	<u>0.2262</u>	0.4792	0.5520	<u>0.1610</u>	<u>1.4700</u>	<u>6.375</u>
	CatBoost	<u>0.9639</u>	0.6242	<u>0.8620</u>	0.2292	0.4792	0.5495	<u>0.1610</u>	<u>1.4654</u>	<u>4.375</u>
	LightGBM	<u>0.9631</u>	0.6164	<u>0.8599</u>	<u>0.2260</u>	0.4877	0.5500	0.1612	<u>1.4654</u>	7.375
	RandomForest	0.9580	0.6068	0.8171	0.2400	0.4841	0.5588	0.1647	1.5845	17.375
	Deep Methods									
	MLP	0.9360	0.6220	0.5508	0.2641	0.4821	0.5515	0.1619	1.5362	16.000
	MLP-PLR	0.9596	0.6185	0.8166	0.2361	0.4799	<u>0.5481</u>	0.1616	1.5235	10.875
	FT-T	0.9591	0.6159	0.5746	0.2438	0.4807	0.5503	0.1623	1.5146	15.125
	TabR	0.9605	0.6148	0.8342	0.2370	0.4883	0.5550	0.1623	1.4732	13.750
	ModernNCA	0.9610	<u>0.6341</u>	0.8378	0.2325	0.4804	0.5520	0.1619	1.4857	9.125
	TabM	<u>0.9640</u>	<u>0.6325</u>	0.8290	0.2306	0.4813	0.5500	0.1612	1.4887	7.250
Adaptive Methods	**Deep Methods with Temporal Embedding (Cai & Ye, 2025)**									
	MLP	0.9471	0.6252	0.5519	0.2431	0.4801	0.5518	0.1621	1.5319	14.375
	MLP-PLR	0.9607	0.6110	0.8158	0.2338	0.4803	0.5494	0.1617	1.5133	11.250
	FT-T	0.9608	0.6211	0.5563	0.2363	<u>0.4778</u>	0.5503	0.1622	1.5118	11.125
	TabR	0.9606	0.6233	0.8426	0.2392	0.4827	0.5497	0.1627	**1.4620**	10.125
	ModernNCA	0.9620	**0.6356**	0.8316	**0.2255**	0.4791	0.5535	0.1617	1.4903	7.625
	TabM	0.9629	0.6271	0.8363	0.2321	0.4791	<u>0.5488</u>	**0.1609**	1.4812	<u>5.125</u>
	Deep Methods with Temporal Modulation (Ours)									
	MLP	0.9593	0.6230	0.5532	0.2345	<u>0.4782</u>	0.5502	0.1616	1.5179	11.000
	MLP-PLR	0.9591	0.6133	0.8190	0.2340	<u>0.4780</u>	**0.5474**	0.1616	1.5134	10.000
	TabM	**0.9641**	<u>0.6318</u>	<u>0.8457</u>	<u>0.2285</u>	**0.4773**	<u>0.5491</u>	<u>0.1610</u>	1.4794	**3.500**

Table 1. Main results on the TabReD benchmark (Rubachev et al., 2025).

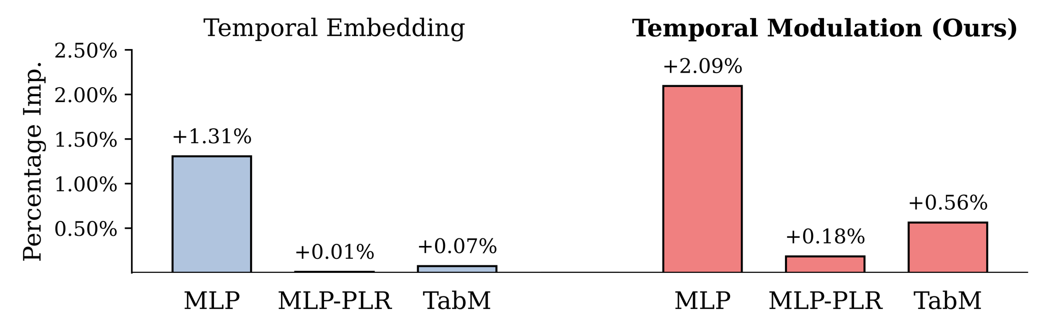

Figure 6. Improvement in relative performance.

In.	Rep.	Out.	Imp. (Relative)	Rank
✗	✗	✗	0.00%	7.000
✗	✗	✓	1.02% (48.7%)	5.125
✗	✓	✗	0.26% (12.6%)	5.250
✗	✓	✓	1.19% (56.8%)	5.500
✓	✗	✗	<u>1.83%</u> (87.4%)	<u>3.250</u>
✓	✗	✓	1.54% (73.6%)	3.625
✓	✓	✗	1.62% (77.2%)	3.750
✓	✓	✓	**2.09%**	**2.500**

Table 2. Ablation study on modulation placement.

Take-away Message

A lightweight, plug-and-play modulation scheme that enables static models to master temporal distribution shifts.

Contact

Paper Code TALENT Survey LinkedIn

References

Cai, H.-R. and Ye, H.-J. Understanding the limits of deep tabular methods with temporal shift. In *ICML*, pp. 6366–6386, 2025.

Perez, E., Strub, F., de Vries, H., Dumoulin, V., and Courville, A. C. Film: Visual reasoning with a general conditioning layer. In *AAAI*, pp. 3942–3951, 2018.

Rubachev, I., Kartashev, N., Gorishniy, Y., and Babenko, A. Tabred: A benchmark of tabular machine learning in-the-wild. In *ICLR*, 2025.