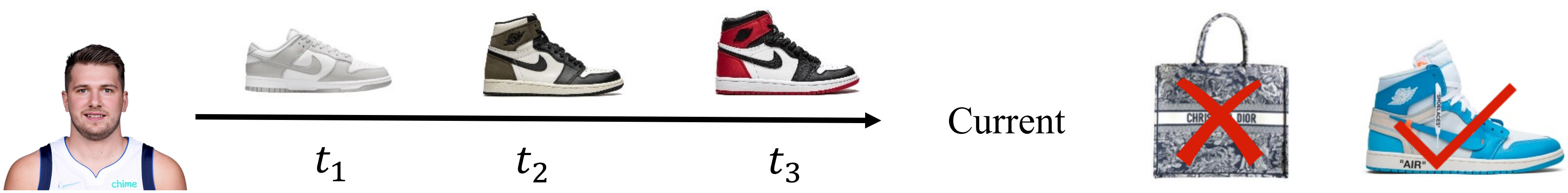
# Temporal Interest Network for User Response Prediction


- Authors: Haolin Zhou* (SJTU), **Junwei Pan*** (Tencent),  Xinyi Zhou (SJTU), Xihua Chen (Tencent), Jie Jiang (Tencent),, Xiaofeng Gao (SJTU), Guihai Chen (SJTU)

# User Response Prediction

## ➢ User Interest Modeling

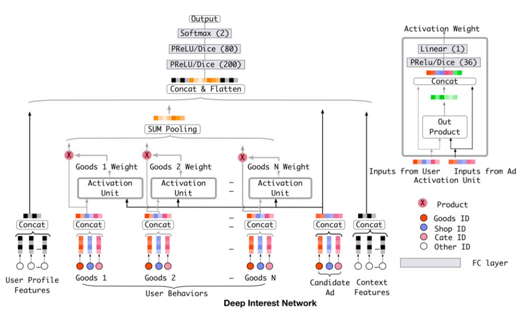
- Recommend items based on his history behavior sequence.



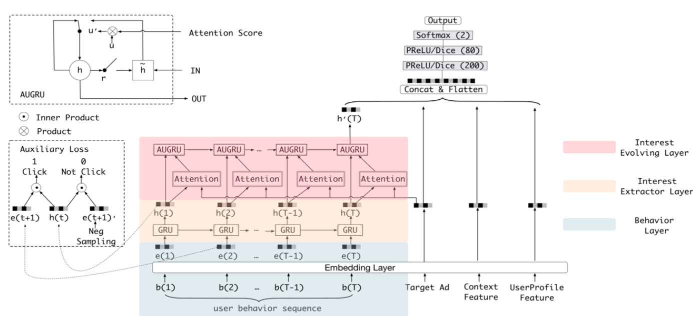

User History Click Sequence

Target

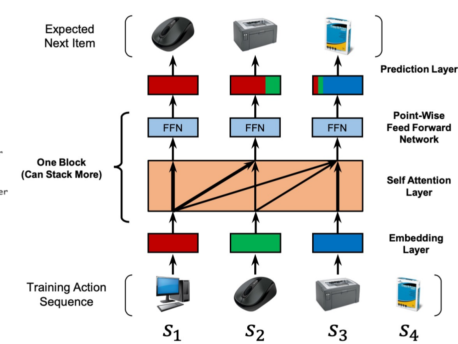
## ➢ Existing Work




DIN



DIEN




SASRec

# Semantic-Temporal Correlation

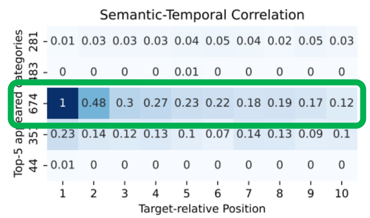
- ➢ Intuitively, there should be both semantic and temporal correlation between behaviors and target. But there is no study on them together.
- ➢ We propose to measure the **ground truth** **semantic-temporal correlation (STC)** by **Mutual Information** between behaviors and the target at the category level.

**Category-wise Target-aware Correlation**

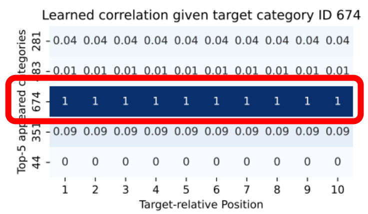
$$Cor = MI(X_{c(X)=c_i \wedge P(x)=p}, Y_{C(Y)=c_t})$$

- ➢ We observe strong Semantic-Temporal Patterns
  - **Semantic** pattern: behaviors belonging to the same category as the target (3rd row in red box) are, in general, more informative than behaviors of the other categories.
  - **Temporal** pattern: there are strong temporal decaying patterns on these behaviors along the position
- ➢ Existing methods, including DIN, SASRec, and BST, fail to learn such patterns.

Target category: 674

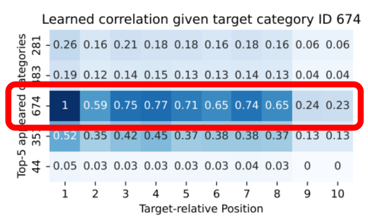



(a) Ground truth STC




(b) DIN's learned STC (0)


(c) SASRec's learned STC (0.8461)




(d) BST's learned STC (0.9894)

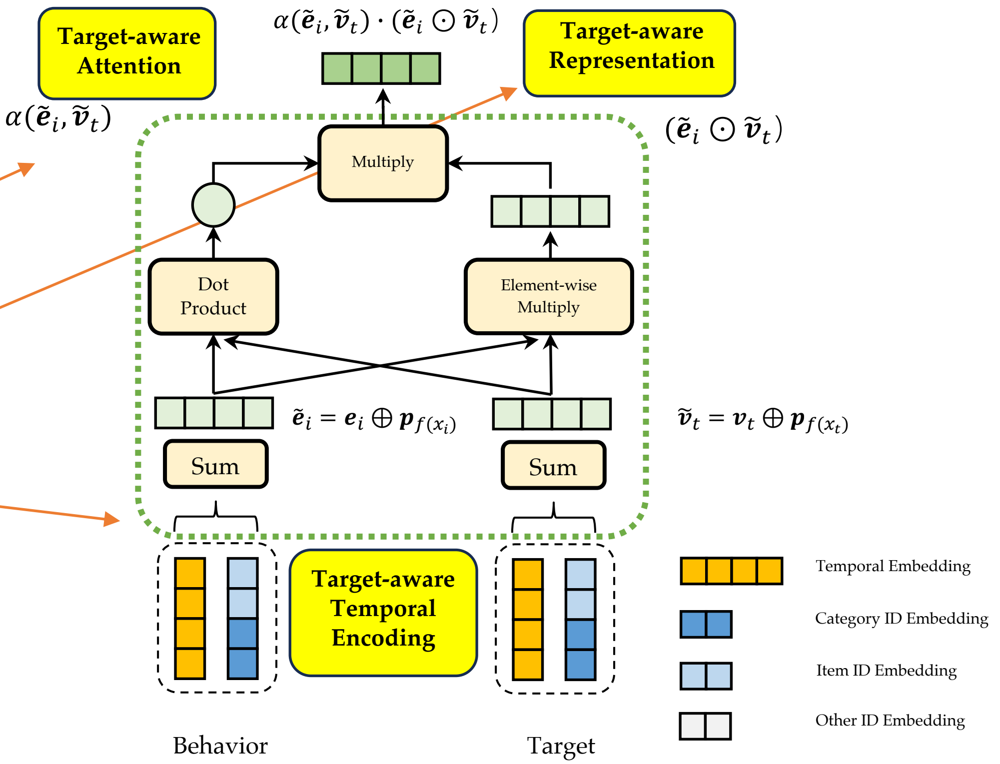
# Temporal Interest Network (TIN)

In order to learn the **semantic-temporal correlation** between behaviors and the target, a model should be able to capture the correlation between **the quadruplet:**

**(behavior semantics, target semantics, behavior temporal, target temporal).**

We propose Temporal Interest Network (TIN). Given a behavior $e_i$ and the target $v_t$, TIN consists of:

- **Target-aware Attention** $\alpha(\tilde{\boldsymbol{e}}_i, \tilde{\boldsymbol{v}}_t)$and **Target-aware Representation** $(\tilde{\boldsymbol{e}}_i \odot \tilde{\boldsymbol{v}}_t)$. We multiply them to conduct **4-way interaction explicitly** between behaviors and the target.
- **Target-aware Temporal Encoding**, which learns a temporal embedding for each behavior according to their relative position or time interval **regarding the target**.



# Performance Evaluation

All three components in TIN are critical.

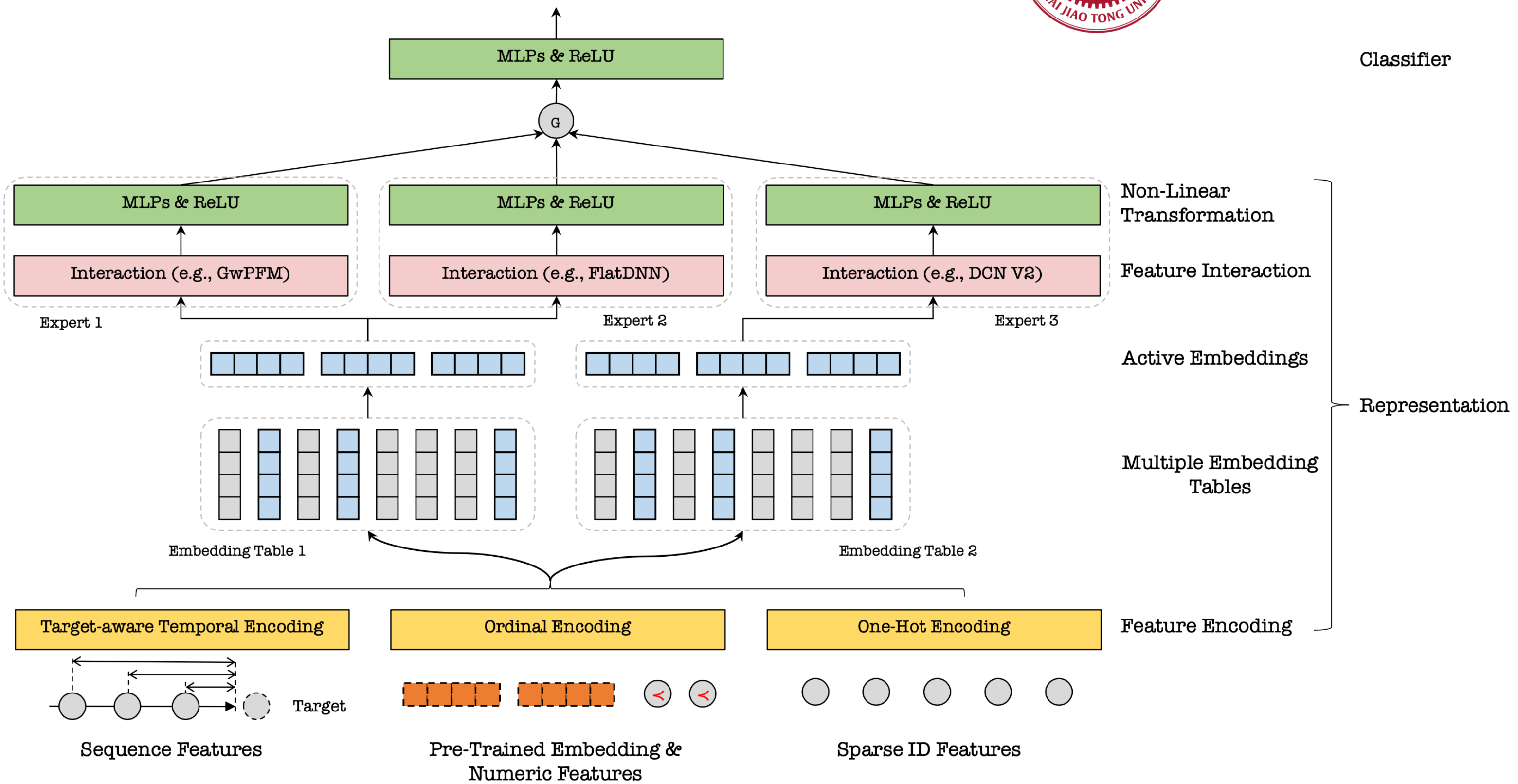
- Without Target-aware Temporal Encoding, the model **loses temporal information** of each behavior regarding the target.
- Without either Target-aware Attention or Target-aware Representation, the model **can't learn the 4-way interaction**.

| Model | Code | Amazon | | | | Alibaba | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Logloss | Δ% | GAUC | Δ% | Logloss | Δ% | GAUC | Δ% |
| Avg Pooling & Concat | ✗✗✗ | 0.4908 (1E-3) | - | 0.8445 (2E-3) | - | 0.1969 (1E-3) | - | 0.6074 (5E-4) | - |
| Avg Pooling & Product | ✗✗✓ | 0.4824 (5E-4) | -1.71 | 0.8523 (8E-4) | +0.92 | 0.1960 (1E-3) | -0.46 | 0.6106 (9E-4) | +0.53 |
| DIN' | ✗✓✗ | 0.4803 (9E-4) | -2.14 | 0.8536 (4E-4) | +1.08 | 0.1962 (1E-3) | -0.36 | 0.6096 (6E-4) | +0.36 |
| DIN | ✗✓✓ | 0.4703 (2E-3) | -4.18 | 0.8590 (1E-3) | +1.72 | 0.1963 (1E-3) | -0.30 | 0.6113 (4E-4) | +0.64 |
| GRU4Rec | ✓✗✓ | 0.4766 (2E-3) | -2.89 | 0.8574 (2E-3) | +1.53 | 0.1972 (1E-3) | +0.15 | 0.6091 (4E-4) | +0.28 |
| SASRec | ✓✗✓ | 0.4837 (7E-3) | -1.45 | 0.8497 (4E-3) | +0.62 | 0.1959 (8E-4) | -0.51 | 0.6091 (4E-4) | +0.28 |
| BERT4Rec | ✓✗✓ | 0.4833 (5E-3) | -1.53 | 0.8501 (2E-3) | +0.66 | 0.1961 (1E-3) | -0.41 | 0.6096 (9E-4) | +0.36 |
| DIEN | ✓✓✗ | 0.4807 (8E-3) | -2.06 | 0.8590 (1E-3) | +1.72 | 0.1973 (9E-4) | +0.20 | 0.6108 (6E-4) | +0.56 |
| DSIN | ✓✓✗ | 0.4726 (2E-3) | -3.71 | 0.8592 (1E-3) | +1.74 | 0.1964 (2E-3) | -0.25 | 0.6106 (9E-4) | +0.53 |
| BST | ✓✓✗ | 0.4850 (5E-4) | -1.18 | 0.8500 (9E-4) | +0.65 | 0.1959 (2E-3) | -0.51 | 0.6096 (6E-4) | +0.36 |
| TIN | ✓✓✓ | **0.4636** (3E-3) | -5.54 | **0.8629** (9E-4) | +2.18 | **0.1954** (2E-3) | -0.76 | **0.6144** (4E-4) | +1.15 |
| TIN w/o TTE | ✗✓✓ | 0.4752 (2E-3) | -3.18 | 0.8544 (8E-4) | +1.17 | 0.1963 (1E-3) | -0.30 | 0.6135 (7E-4) | +1.00 |
| TIN w/o TA | ✓✗✓ | 0.4758 (3E-3) | -3.06 | 0.8566 (1E-3) | +1.43 | 0.1960 (7E-4) | -0.46 | 0.6094 (8E-4) | +0.33 |
| TIN w/o TR | ✓✓✗ | 0.4743 (2E-3) | -3.36 | 0.8576 (9E-4) | +1.55 | 0.1965 (2E-3) | -0.20 | 0.6127 (1E-3) | +0.87 |

TIN outperforms the AUC of best-performing baselines by **0.43%** and **0.29%** on two public datasets.

Code: denoting which of the three components (Temporal Information, Target-aware Attention and Target-aware Representation) each method has.
Models with the same background color have the same code.

# Online Deployment at Tencent




- Production model: MoE architecture [1], consisting of feature encoding and feature interaction experts.
- Add TIN as an additional expert on the sequence features.

## ➢ Online A/B Testing

- During online A/B testing in pCTR prediction of WeChat Moments, TIN achieved a **1.65% cost lift,** and a **1.93% GMV (Gross Merchandise Value) lift** over the base model.
- TIN has been successfully deployed to pCTR and pLTV prediction tasks of many scenarios at Tencent.

[1]. Ad Recommendation in a Collapsed and Entangled World. Under Review, Tencent.

# Visualization

➢ Measurement of Learned Semantic Temporal Correlation of all Models

- Definition of the learned correlation of a given model:

$$\overline{Cor}(f(X_i), e_i, v_t) = e^z \cdot \|r\|_2$$

- For TIN, which is formulated as

$$\boldsymbol{u}_{\text{TIN}} = \sum_{i \in \mathcal{H}} \alpha(\tilde{\boldsymbol{e}}_i, \tilde{\boldsymbol{v}}_t) \cdot (\tilde{\boldsymbol{e}}_i \odot \tilde{\boldsymbol{v}}_t)$$

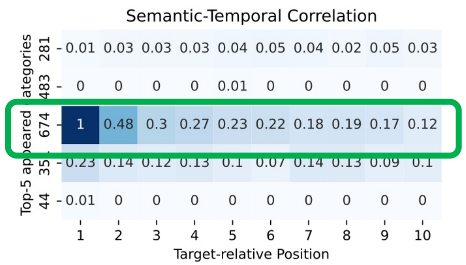
Its learned correlation can be measured by:

$$\boxed{e^{\langle \tilde{\boldsymbol{e}}_i, \tilde{\boldsymbol{v}}_t \rangle}} \cdot \boxed{\|\tilde{\boldsymbol{e}}_i \odot \tilde{\boldsymbol{v}}_t\|_2}$$

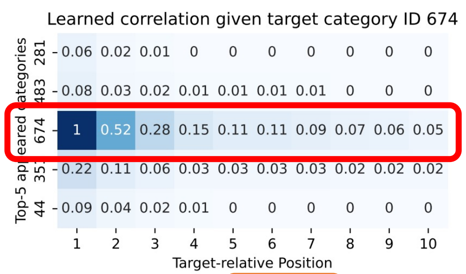
Attention:
Dot Product

Representation:
Norm of Embedding

Target category: 674

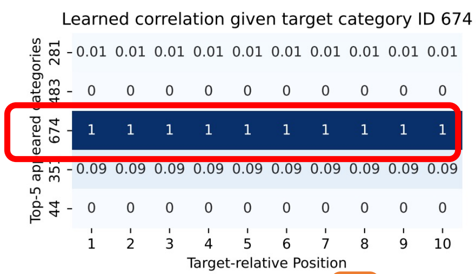



(a) Ground truth STC

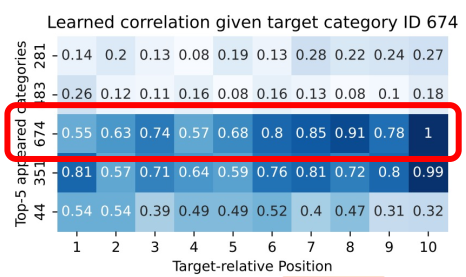



(a) TIN (0.9894)

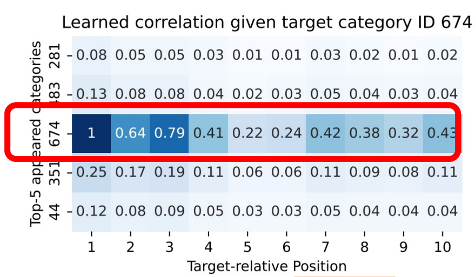
Pearson correlation coefficient to the ground truth




(b) TIN w/o TTE (0)




(c) TIN w/o TA (-0.6845)




(d) TIN w/o TR (0.8179)

# Connections to Existing Methods

$$\boldsymbol{u}_{\text{TIN}} = \sum_{i \in \mathcal{H}} \underbrace{\overbrace{\alpha(\tilde{\boldsymbol{e}}_i, \tilde{\boldsymbol{v}}_t)}^{\text{TA}} \cdot \overbrace{(\tilde{\boldsymbol{e}}_i \odot \tilde{\boldsymbol{v}}_t)}^{\text{TR}}}_{\text{TTE}}$$

- Summary of existing methods regarding the existence of three components: Temporal Information (TI), Target-aware Attention (TA), and Target-aware Representation (TR).
- **All existing methods lack one or several of these three components**, making them **fail to capture the 4-way** semantic-temporal correlation.

| Model | Temporal Information? | How | TA | Attention | TR | Representation |
|---|---|---|---|---|---|---|
| Avg Pooling & Concat | ✗ | - | ✗ | - | ✗ | $g_{\text{MLP}}([\boldsymbol{e}_i, \boldsymbol{v}_t])$ |
| Avg Pooling & Product | ✗ | - | ✗ | - | ✓ | $g_{\text{MLP}}([\boldsymbol{e}_i, \boldsymbol{v}_t, \boldsymbol{e}_i \odot \boldsymbol{v}_t])$ |
| DIN' [45] | ✗ | - | ✓ | $\sigma(\langle \boldsymbol{e}_i, \boldsymbol{v}_t \rangle)$ | ✗ | $g_{\text{MLP}}([\boldsymbol{e}_i, \boldsymbol{v}_t])$ |
| DIN [45] | ✗ | - | ✓ | $\sigma(\langle \boldsymbol{e}_i, \boldsymbol{v}_t \rangle)$ | ✓ | $g_{\text{MLP}}([\boldsymbol{e}_i, \boldsymbol{v}_t, \boldsymbol{e}_i \odot \boldsymbol{v}_t])$ |
| SIM [21] | ✓ | TTE-T | ✓ | $\sigma(\langle \boldsymbol{e}_i, \boldsymbol{v}_t \rangle)$ | ✗ | $g_{\text{MLP}}([\boldsymbol{e}_i, \boldsymbol{v}_t])$ |
| GRU4Rec [7] | ✓ | GRU | ✗ | - | ✓ | $\langle \boldsymbol{h}_i, \boldsymbol{v}_t \rangle$ |
| SASRec [10] | ✓ | COE | ✗ | - | ✓ | $\langle \tilde{\boldsymbol{e}}_i^V, \boldsymbol{v}_t \rangle$ |
| BERT4Rec [29] | ✓ | COE | ✗ | - | ✓ | $\langle \tilde{\boldsymbol{e}}_i^V, \boldsymbol{v}_t \rangle$ |
| ExtLayer [44] | ✓ | AUGRU | ✗ | - | ✗ | - |
| EvoLayer [44] | ✓ | AUGRU | ✓ | $\sigma(\langle \boldsymbol{h}_i, \boldsymbol{v}_t \rangle)$ | ✗ | $g_{\text{MLP}}([\boldsymbol{h}_i, \boldsymbol{v}_t])$ |
| DSIN [5] | ✓ | Bi-LSTM & COE | ✓ | $\sigma(\langle \tilde{\boldsymbol{e}}_i^K, \boldsymbol{v}_t^Q \rangle)$ | ✗ | $g_{\text{MLP}}([\tilde{\boldsymbol{e}}_i^V, \boldsymbol{v}_t])$ |
| BST [3] | ✓ | COE | ✓ | $\sigma(\langle \tilde{\boldsymbol{e}}_i^K, \tilde{\boldsymbol{v}}_t^Q \rangle)$ | ✗ | $g_{\text{MLP}}([\tilde{\boldsymbol{e}}_i^V, \boldsymbol{v}_t])$ |
| TIN | ✓ | TTE | ✓ | $\sigma(\langle \tilde{\boldsymbol{e}}_i, \tilde{\boldsymbol{v}}_t \rangle)$ | ✓ | $g_{\text{MLP}}([\tilde{\boldsymbol{e}}_i, \tilde{\boldsymbol{v}}_T, \tilde{\boldsymbol{e}}_i \odot \tilde{\boldsymbol{v}}_t])$ |
| TIN w/o TTE | ✗ | - | ✓ | $\sigma(\langle \tilde{\boldsymbol{e}}_i, \tilde{\boldsymbol{v}}_t \rangle)$ | ✓ | $g_{\text{MLP}}([\tilde{\boldsymbol{e}}_i, \tilde{\boldsymbol{v}}_t, \tilde{\boldsymbol{e}}_i \odot \tilde{\boldsymbol{v}}_T])$ |
| TIN w/o TA | ✓ | TTE | ✗ | - | ✓ | $g_{\text{MLP}}([\tilde{\boldsymbol{e}}_i, \tilde{\boldsymbol{v}}_t, \tilde{\boldsymbol{e}}_i \odot \tilde{\boldsymbol{v}}_t])$ |
| TIN w/o TR | ✓ | TTE | ✓ | $\sigma(\langle \tilde{\boldsymbol{e}}_i, \tilde{\boldsymbol{v}}_t \rangle)$ | ✗ | $g_{\text{MLP}}([\tilde{\boldsymbol{e}}_i, \tilde{\boldsymbol{v}}_t])$ |

# Summary

- We propose to measure **the semantic-temporal correlation** between behaviors and the target by **Mutual Information** and observe **strong semantic and temporal patterns** simultaneously.
- We find that all existing models **fail to** capture such strong semantic-temporal patterns.
- We propose TIN, which consists of three key components: **Target-aware Temporal Encoding, Target-aware Attention, and Target-aware Representation**. All existing methods **lack one or several** of these three components.
- We demonstrated TIN's effectiveness on public datasets and deployed it on Tencent’s Ads Platform.
- By measuring the learned correlation, we confirm TIN **captures the semantic-temporal pattern** decently.

# Q & A

Code: https://github.com/zhouxy1003/TIN
Contact: jonaspan@tencent.com

Github QR Code