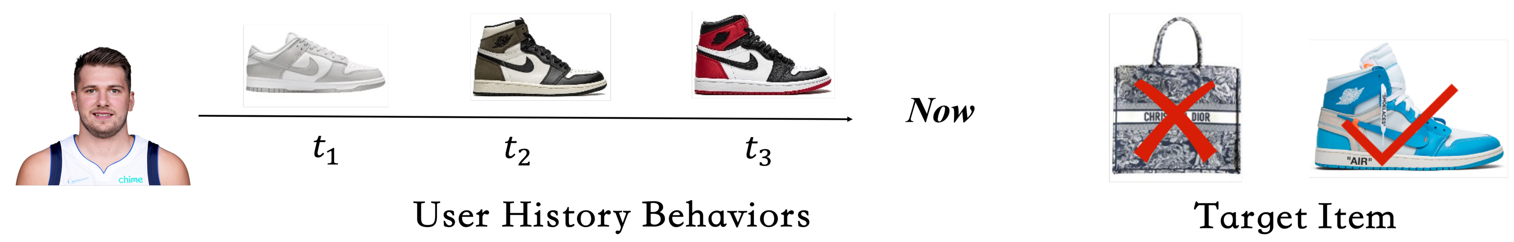
Temporal Interest Network for User Response Prediction

Haolin Zhou[1*], **Junwei Pan**[2*], Xinyi Zhou[1], Xihua Chen[2], Jie Jiang[2], Xiaofeng Gao[1], Guihai Chen[1],
[1]Shanghai Jiao Tong University [2]Tencent Inc.

User Interest Modeling

User interest modeling based on her history behaviors has been a hot research topic in recommender systems. There should be both semantic and temporal correlation between behaviors and the target.

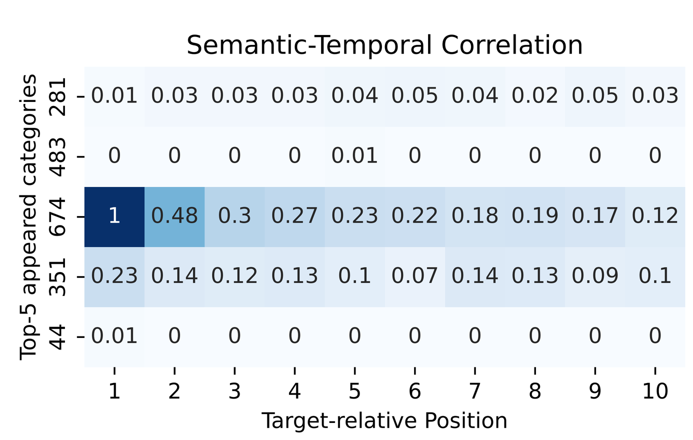


How do Existing Method Learn Semantic-Temporal Correlation?

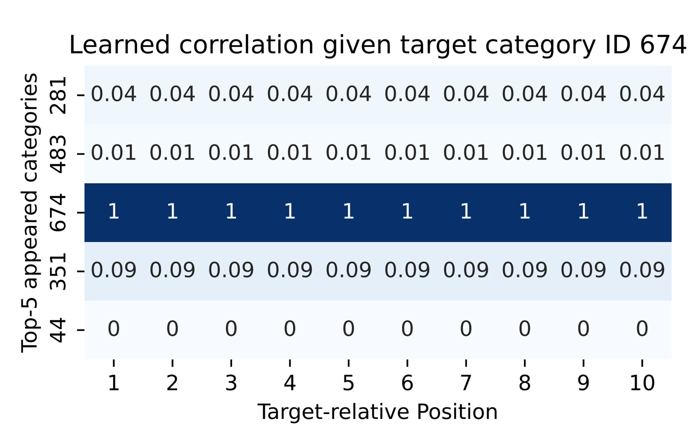
Definition (**Category-wise Target-aware Correlation** (CTC)). CTC is defined as the mutual information between behaviors with category c_i whilst occurring at position p, and the user response label on the target item with category c_t. Formally,

$$\text{Cor} = \text{MI}\left(\mathcal{X}_{C(X)=c_i \wedge P(X)=p}, \mathcal{Y}_{C(Y)=c_t}\right), \quad (1)$$

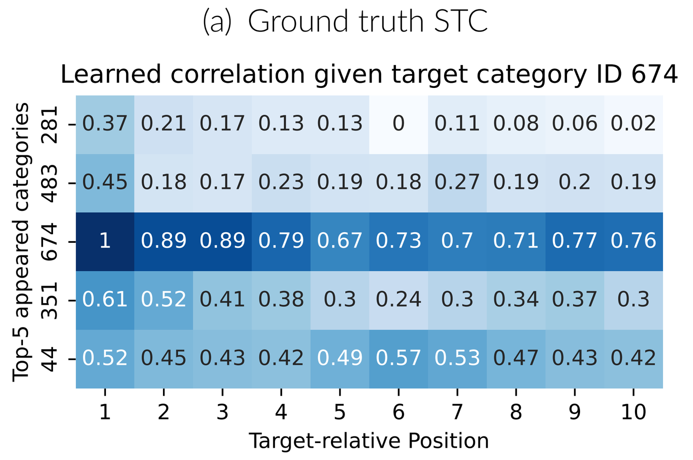
where $C(\cdot)$, $P(\cdot)$ denotes the category or position of the behavior or target.



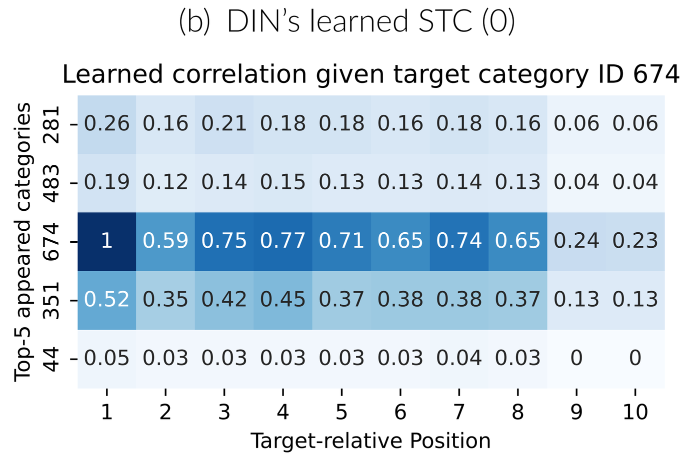
(a) Ground truth STC



(b) DIN's learned STC (0)



(c) SASRec's learned STC (0.8461)



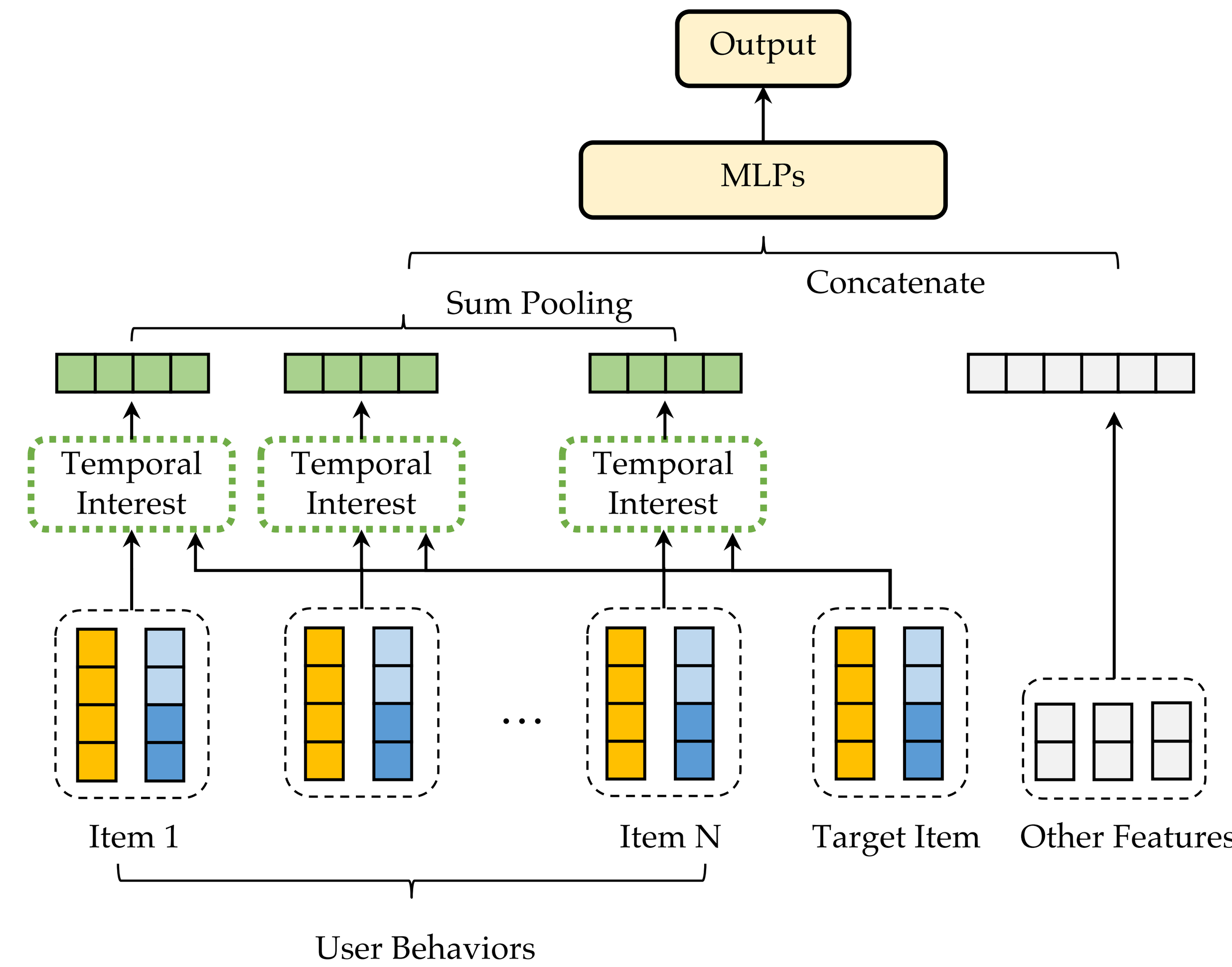
(d) BST's learned STC (0.6245)

We observe **strong semantic-temporal correlations** between behaviors and the target:

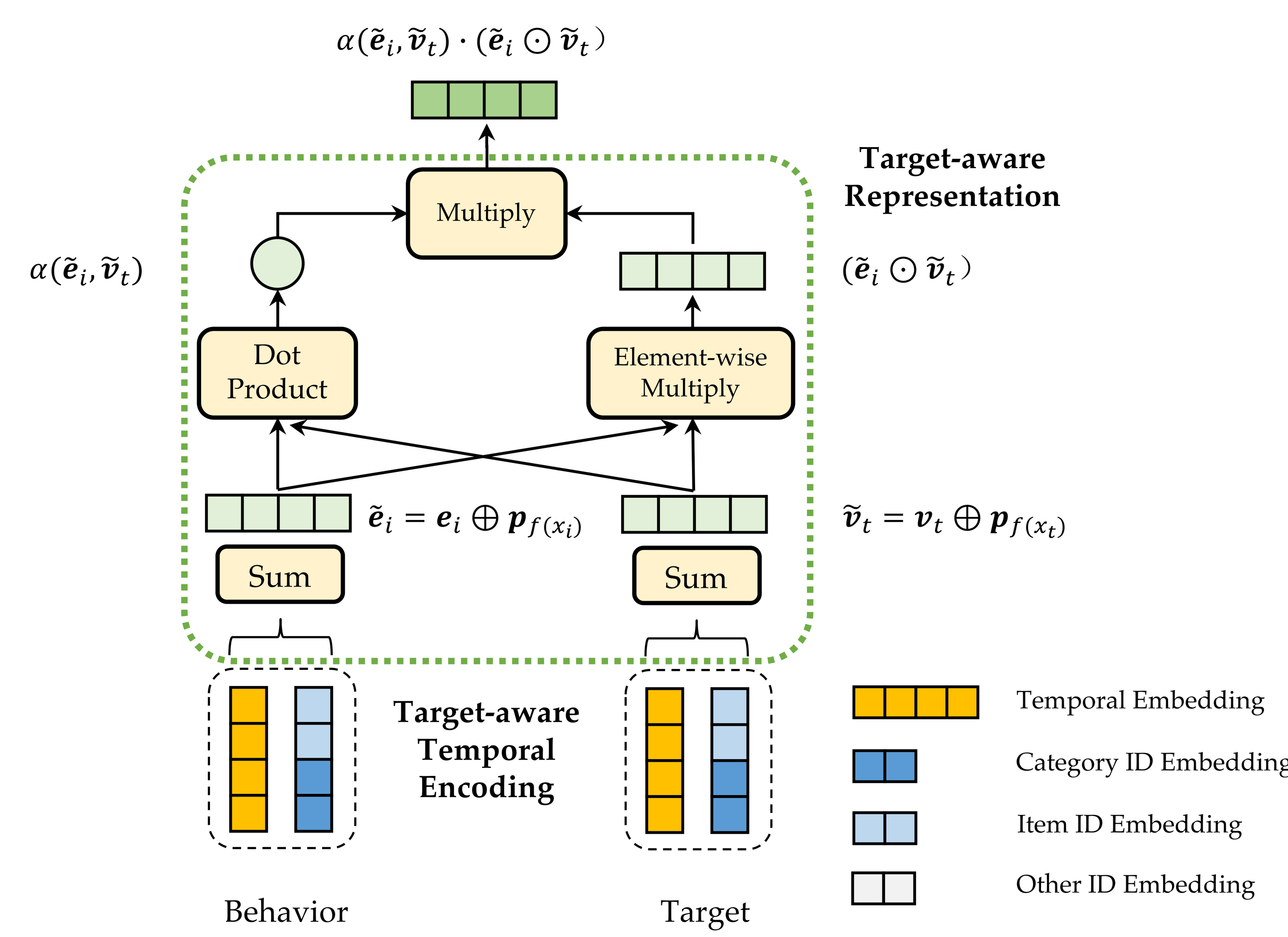
semantic pattern between matching categories: behaviors belonging to the same category as the target exhibit a higher degree of correlation than other categories.

temporal decaying pattern: among the semantically correlated behaviors, there is a compelling correlation decrease from the most recent behaviors to the oldest ones.

Temporal Interest Network



Temporal Interest Module



In order to conduct 4-way interaction over the quadruplet `(behavior semantics, target semantics, behavior temporal, target temporal)` to capture the semantic-temporal correlation, we propose TIM, which consists of a) **Target-aware Temporal Encoding (TTE)**: TTE preserves the temporal information of behaviors regarding the target. b) **Target-aware Attention (TA)** and c) **Target-aware Representation (TR)** over the temporally encoded behaviors and target.

$$\boldsymbol{u}_{\text{TIN}} = \sum_{i \in \mathcal{H}} \underbrace{\overbrace{\alpha(\tilde{\boldsymbol{e}}_i, \tilde{\boldsymbol{v}}_t)}^{\text{TA}} \cdot \overbrace{(\tilde{\boldsymbol{e}}_i \odot \tilde{\boldsymbol{v}}_t)}^{\text{TR}}}_{\text{TTE}} \quad (2)$$

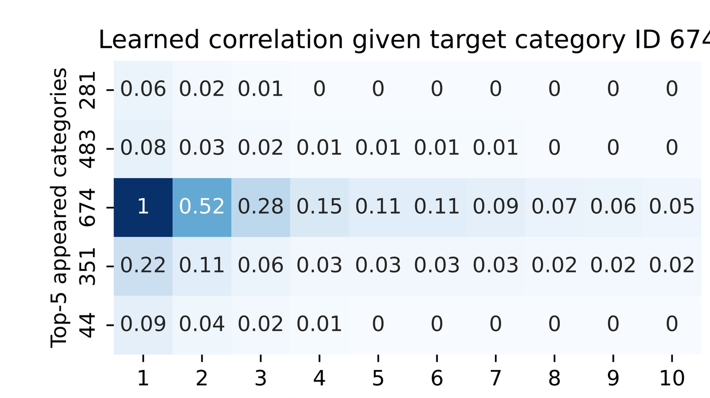
Performance Evaluation

Table 1. Models with the same background color have the same component code.

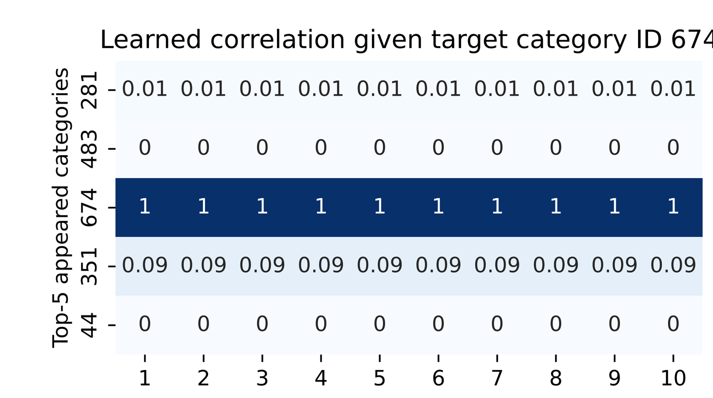
Model	Code	Amazon Logloss	Amazon Δ%	Amazon GAUC	Amazon Δ%	Alibaba Logloss	Alibaba Δ%	Alibaba GAUC	Alibaba Δ%
Avg Pooling & Concat	✗✗✗	0.4908 (1E-3)	-	0.8445 (2E-3)	-	0.1969 (1E-3)	-	0.6074 (5E-4)	-
Avg Pooling & Product	✗✗✓	0.4824 (5E-4)	-1.71	0.8523 (8E-4)	+0.92	0.1960 (1E-3)	-0.46	0.6106 (9E-4)	+0.53
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
	✓✓✓	**0.4636** (3E-3)	-5.54	**0.8629** (9E-4)	+2.18	**0.1954** (2E-3)	-0.76	**0.6144** (4E-4)	+1.15
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
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
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 has been successfully deployed to the pCTR and pLTV models on many Tencent Ads scenarios, including Wechat Moments, Channel, Tencent Video, Tencent News.

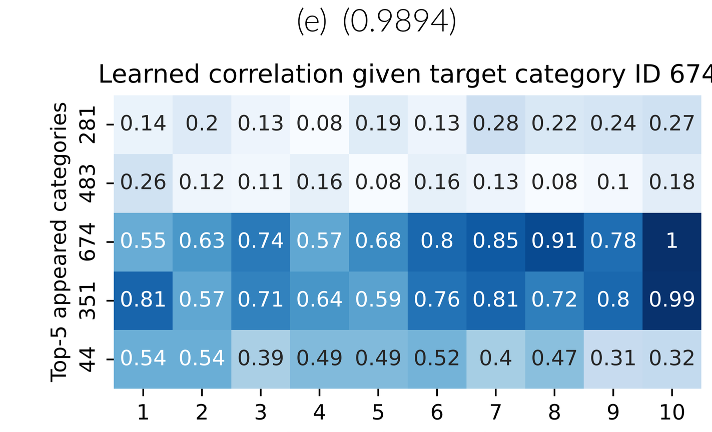
TIN Captures Semantic-Temporal Correlation Decently



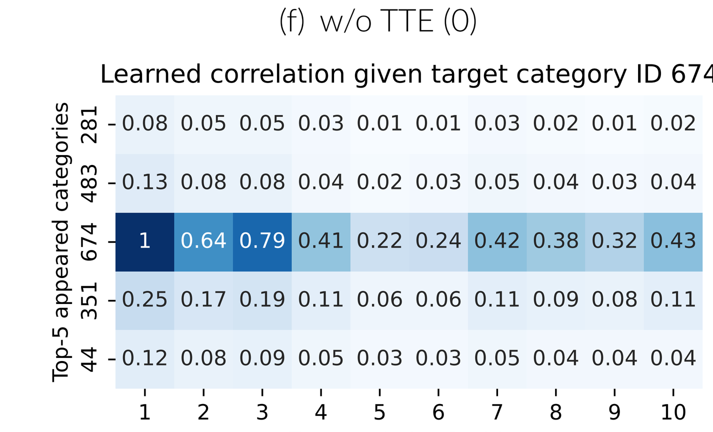
(e) (0.9894)



(f) w/o TTE (0)



(g) w/o TA (-0.6845)



(h) w/o TR (0.8179)

Connections to Existing Methods

Model	Temporal Information?	How	TA	Attention	TR	Representation	Notes on 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'	✗	-	✓	$\sigma(\langle \boldsymbol{e}_i, \boldsymbol{v}_t \rangle)$	✗	$g_{\text{MLP}}([\boldsymbol{e}_i, \boldsymbol{v}_t])$	Concat & MLP's hard to learn dot product ?.
DIN	✗	-	✓	$\sigma(\langle \boldsymbol{e}_i, \boldsymbol{v}_t \rangle)$	✓	$g_{\text{MLP}}([\boldsymbol{e}_i, \boldsymbol{v}_t, \boldsymbol{e}_i \odot \boldsymbol{v}_t])$	According to code released by the authors.
SIM	✓	TTE-T	✓	$\sigma(\langle \boldsymbol{e}_i, \boldsymbol{v}_t \rangle)$	✗	$g_{\text{MLP}}([\boldsymbol{e}_i, \boldsymbol{v}_t])$	Concat & MLP's hard to learn dot product ?.
GRU4Rec	✓	GRU	✗	-	✓	$\langle \boldsymbol{h}_i, \boldsymbol{v}_t \rangle$	According to open source code.
SASRec	✓	COE	✗	-	✓	$\langle \tilde{\boldsymbol{e}}_i^V, \boldsymbol{v}_t \rangle$	According to Eqn. $r_{t,i} = \boldsymbol{F}_t^{(b)} \boldsymbol{N}_i^T$ of ?.
BERT4Rec	✓	COE	✗	-	✓	$\langle \tilde{\boldsymbol{e}}_i^V, \boldsymbol{v}_t \rangle$	According to Eqn.(7) in ?.
ExtLayer	✓	AUGRU	✗	-	✗	-	-
EvoLayer	✓	AUGRU	✓	$\sigma(\langle \boldsymbol{h}_i, \boldsymbol{v}_t \rangle)$	✗	$g_{\text{MLP}}([\boldsymbol{h}_i, \boldsymbol{v}_t])$	-
DSIN	✓	Bi-LSTM & COE	✓	$\sigma(\langle \tilde{\boldsymbol{e}}_i^K, \boldsymbol{v}_t^Q \rangle)$	✗	$g_{\text{MLP}}([\tilde{\boldsymbol{e}}_i^V, \boldsymbol{v}_t])$	Concat & MLP's hard to learn dot product ?.
BST	✓	COE	✓	$\sigma(\langle \tilde{\boldsymbol{e}}_i^K, \tilde{\boldsymbol{v}}_t^Q \rangle)$	✗	$g_{\text{MLP}}([\tilde{\boldsymbol{e}}_i^V, \boldsymbol{v}_t])$	Concat & MLP's hard to learn dot product ?.
	✓	TTE	✓	$\sigma(\langle \tilde{\boldsymbol{e}}_i, \tilde{\boldsymbol{v}}_t \rangle)$	✓	$g_{\text{MLP}}([\tilde{\boldsymbol{e}}_i, \tilde{\boldsymbol{v}}_T, \tilde{\boldsymbol{e}}_i \odot \tilde{\boldsymbol{v}}_t])$	-
w/o TTE	✗	-	✓	$\sigma(\langle \tilde{\boldsymbol{e}}_i, \tilde{\boldsymbol{v}}_t \rangle)$	✓	$g_{\text{MLP}}([\tilde{\boldsymbol{e}}_i, \tilde{\boldsymbol{v}}_t, \tilde{\boldsymbol{e}}_i \odot \tilde{\boldsymbol{v}}_T])$	-
w/o TA	✓	TTE	✗	-	✓	$g_{\text{MLP}}([\tilde{\boldsymbol{e}}_i, \tilde{\boldsymbol{v}}_t, \tilde{\boldsymbol{e}}_i \odot \tilde{\boldsymbol{v}}_t])$	-
w/o TR	✓	TTE	✓	$\sigma(\langle \tilde{\boldsymbol{e}}_i, \tilde{\boldsymbol{v}}_t \rangle)$	✗	$g_{\text{MLP}}([\tilde{\boldsymbol{e}}_i, \tilde{\boldsymbol{v}}_t])$	-

All existing methods lack one or several of the three components (Temporal Information, Target-aware Attention and Target-aware Representation), making them fail to learn the semantic-temporal correlation.